Econometric Theory, **38**, 2022, 689–751.
doi:10.1017/S026646662100027X

IDENTIFICATION ROBUST INFERENCE FOR MOMENTS-BASED ANALYSIS OF LINEAR DYNAMIC PANEL DATA MODELS

MAURICE J.G. BUN
De Nederlandsche Bank
University of Amsterdam

FRANK KLEIBERGEN
University of Amsterdam

We use identification robust tests to show that difference (Dif), level (Lev), and nonlinear (NL) moment conditions, as proposed by Arellano and Bond (1991, *Review of Economic Studies* 58, 277–297), Ahn and Schmidt (1995, *Journal of Econometrics* 68, 5–27), Arellano and Bover (1995, *Journal of Econometrics* 68, 29–51), and Blundell and Bond (1998, *Journal of Econometrics* 87, 115–143) for the linear dynamic panel data model, do not separately identify the autoregressive parameter when its true value is close to one and the variance of the initial observations is large. We prove that combinations of these moment conditions, however, do so when there are more than three time series observations. This identification then solely results from a set of, so-called, robust moment conditions. These robust moments are spanned by the combined Dif, Lev, and NL moment conditions and only depend on differenced data. We show that, when only the robust moments contain identifying information on the autoregressive parameter, the discriminatory power of the Kleibergen (2005, *Econometrica* 73, 1103–1124) Lagrange multiplier (KLM) test using the combined moments is identical to the largest rejection frequencies that can be obtained from solely using the robust moments. This shows that the KLM test implicitly uses the robust moments when only they contain information on the autoregressive parameter.

1. INTRODUCTION

It is common to estimate the parameters of linear dynamic panel data models using the generalized method of moments (GMM; Hansen, 1982). The moment conditions for the linear dynamic panel data model either analyze it in first

The research of the first author has been funded by the NWO Vernieuwingsimpuls research grant "Causal Inference with Panel Data." We thank the Editor, Peter Phillips, the Co-Editor, Guido Kuersteiner, two anonymous referees, Manuel Arellano, Richard Blundell, Steve Bond, Peter Boswijk, Geert Dhaene, Frank Windmeijer, and participants of seminars at Bristol, CEMFI, and CORE, the Cowles Summer Conference at Yale, the EC^2 Meeting in Maastricht, Groningen, Leuven, and Monash, and the 19th International Conference on Panel Data in London and Oxford, the Tinbergen Institute in Amsterdam and Toulouse, and UCL for helpful comments and discussion. Address correspondence to Frank Kleibergen, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; e-mail: f.r.kleibergen@uva.nl.

This article was originally published with a missing reference. A notice detailing this has been published, and the reference added to the online PDF and HTML versions.

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

differences using lagged levels of the series as instruments, in levels using lagged first differences as instruments or using a product of levels and first differences. We refer to the first set of moment conditions as Dif(ference) moment conditions (see Arellano and Bond, 1991), the second set as Lev(el) moment conditions (see Arellano and Bover, 1995; Blundell and Bond, 1998), and the third set as N(on-)L(inear) moment conditions (see Ahn and Schmidt, 1995).

The Dif, Lev, and NL moment conditions can be used separately to identify the parameters of dynamic panel data models. To exhaust all information, however, two particular combinations of Dif, Lev, and NL moment conditions have been proposed. We refer to the combined Dif and Lev moment conditions as the Sys(tem) moment conditions and the combination of the Dif and NL moment conditions as the A(hn–)S(chmidt) moment conditions.[1] The Sys moment conditions exhaust all information on the autoregressive parameter that is present under mean stationarity (see Arellano and Bover, 1995; Blundell and Bond, 1998). The AS moment conditions exhaust all information while not assuming mean stationarity (see Ahn and Schmidt, 1995).

We analyze the identification of the autoregressive parameter by the various sets of moment conditions for a range of true values including the case of highly persistent panel data. All moment conditions involve first differences of the series to remove individual specific effects. The first difference operator removes information in the time series at the unit root value of the autoregressive parameter. It is well known that the Dif moment conditions, therefore, do not identify the autoregressive parameter when its true value is (close to) one, since lagged levels are then weak predictors of first differences. This has led to the development of the NL and Lev, and hence AS and Sys, moment conditions which were originally considered to identify the autoregressive parameter when the panel data are highly persistent.

To show the identification issues at specific values of the autoregressive parameter, we use identification robust tests, i.e., the GMM–A(nderson–)R(ubin) statistic of Anderson and Rubin (1949) and Stock and Wright (2000), and the K(leibergen) L(agrange) M(ultiplier) statistic of Kleibergen (2005). At values of the parameters where identification issues occur, the rejection frequency of these tests provenly coincides with the significance level, so the identification issues are relatively easy to detect by inspecting the power curves. Using power curves of the KLM test, we show that Dif, Lev, and NL moment conditions separately do not identify the autoregressive parameter for persistent values of it when paired with a large variance of the initial observations. The same holds for the Sys moment conditions with three times series observations. The power curves further show that Sys and AS moment conditions generally identify the autoregressive parameter when the number of time series observations exceeds 3.

We formally prove these identification results using an asymptotic sampling scheme in which we jointly let the variance of the initial observations and the

[1] Note that in a combination of all three sets of moments conditions, the NL moment conditions are redundant.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

number of cross section observations go to infinity. For a range of relative convergence rates of the variance of the initial observations compared to the cross section sample size, the Dif, Lev, and NL sample moments and their derivatives diverge. Both the population moment and the Jacobian identification condition are then ill defined, which implies that the autoregressive parameter is not separately identified by the Dif, NL, or Lev moment conditions. These results confirm and extend earlier findings in Madsen (2003), Bond, Nauges, and Windmeijer (2005), Hahn, Hausman, and Kuersteiner (2007), Kruiniger (2009), and Phillips (2018).

Using our asymptotic sampling scheme, we also prove that AS and Sys moment conditions identify the autoregressive parameter irrespective of the variance of the initial observation when the number of time series observations exceeds 3. When the variance of the initial observations is large, the identification results from a set of, so-called, robust sample moments that are a combination of the Dif, Lev, and NL sample moments (other than AS and Sys) and only depend on differenced data. These robust sample moments are spanned by the Sys sample moments and also by the AS sample moments. They identify the autoregressive parameter irrespective of the variance of the initial observation and including the case of highly persistent data. They are a subset of the moment conditions in Kruiniger (2002), which are derived under the additional assumption of time series homoskedasticity.

Despite these positive identification results for the Sys and AS moments, the large sample distributions of corresponding one-step and two-step GMM estimators are known to be nonstandard when the variance of the initial observation is large and the autoregressive parameter is close to one. This makes it hard to infer if and how standard GMM inference using the original AS or Sys sample moments exploits the information contained in the robust sample moments that they encompass. The nonstandard limiting behavior results, since the identification of the autoregressive parameter is then of, so-called, second order, since the Jacobian of the robust sample moments is rank deficient, but the Hessian is not (see, e.g., Dovonon and Renault, 2013; Dovonon and Hall, 2018; Dovonon, Hall, and Kleibergen, 2020). It explains the large biases of the one-step and two-step GMM estimators and the size distortions of their corresponding t-statistics when the series are persistent (see, e.g., Madsen, 2003; Bond et al., 2005; Bond and Windmeijer, 2005; Hahn et al., 2007; Kruiniger, 2009; Bun and Windmeijer, 2010; Dhaene and Jochmans, 2016). Because of the second-order identification, GMM estimators based on the robust sample moments also have nonstandard asymptotic distributions when the data are persistent (see Dovonon et al., 2020).

We therefore analyze how identification robust test statistics exploit the identifying information in the robust sample moments. We prove that the identification robust KLM test procedure based on either AS or Sys sample moments exploits all the identifying information contained in the robust sample moments. We do so by first determining the (infeasible) optimal weighted average of the robust sample moments that maximizes the discriminatory power of a GMM-AR test of the autoregressive parameter in settings where only the robust sample moments contain identifying information. Next, we determine the discriminatory power of

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

KLM tests, based on AS or Sys moment conditions, under such settings and prove that it equals that of the GMM-AR test using the optimal weighted average of the robust sample moments. KLM tests using AS or Sys moment conditions thus resort to just using the robust sample moments when only the latter contain information on the autoregressive parameter. It is therefore not necessary to explicitly use the robust sample moments, which provide identification under mild conditions, since they are implicitly used in the KLM test based on either AS or Sys sample moments.

The paper is organized as follows. Section 2 introduces the linear dynamic panel data model and the different moment conditions we use to identify its parameters. It also discusses identification robust statistics, specifically the KLM test, that we use to illustrate the identification issues that occur at persistent values of the autoregressive parameter. In Section 3, we use a representation theorem, akin to the cointegration representation theorem (see Engle and Granger, 1987; Johansen, 1991) to pin down the identification properties of the different moment conditions. This theorem also allows us to obtain the robust sample moments. In Section 4, we define the GMM-AR test that uses the (infeasible) optimal weighted average of the robust sample moments and derive the large sample distribution of the KLM test using AS or Sys moment conditions under settings where only the robust sample moments contain information on the autoregressive parameter. The fifth (final) section concludes. Proofs of theorems and definitions of sample moments are provided in the Appendix. We use the following notation throughout the paper: vec(A) stands for the (column) vectorization of the $k \times n$ matrix A, vec$(A) = (a_1' \dots a_n')'$, for $A = (a_1 \dots a_n), P_A = A(A'A)^{-1}A'$ is a projection on the columns of the full-rank matrix A, and $M_A = I_N - P_A$ is a projection on the space orthogonal to A. Convergence in probability is denoted by "$\underset{p}{\rightarrow}$", convergence in distribution by "$\underset{d}{\rightarrow}$", and "$\underset{a}{=}$" means asymptotically equivalent.

2. IDENTIFICATION ROBUST GMM INFERENCE FOR DYNAMIC PANEL DATA MODELS

In this section, we briefly describe the dynamic panel data model and the different sets of moment conditions. Thereafter, we discuss identification robust GMM inference including the construction of confidence intervals. Finally, we illustrate the identification issues that occur when using the different moment conditions for dynamic panel data models, by computing power curves based on the identification robust KLM statistic.

2.1. Model and Moment Conditions

We analyze the first-order autoregressive linear dynamic panel data model

$$y_{it} = c_i + \theta y_{it-1} + u_{it}, \quad i = 1, \dots, N, t = 2, \dots, T, \tag{1}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

with T the number of time periods and N the number of cross section observations. We assume that the initial observation y_{i1} is observed and that the vector of observations $(y_{i1}, \ldots, y_{iT})$ for individual i is independently distributed across the N individuals. We will later on make further assumptions on the initial observations to properly define the process in (1). For expository purposes, we analyze the simple dynamic panel data model in (1), which can be extended with additional lags of y_{it} and explanatory variables.[2] Estimation of the parameter θ by means of least squares leads to an inconsistent estimator in samples with a finite value of T and large N (see, e.g., Nickell, 1981). We therefore estimate it using GMM. We obtain the GMM moment conditions from the unconditional moment assumptions:

$$\begin{array}{rll} E[u_{it}] & = 0, & t = 2, \ldots, T, \\ E[u_{it}u_{is}] & = 0, & s \neq t; s, t = 2, \ldots, T, \\ E[u_{it}c_i] & = 0, & t = 2, \ldots, T, \\ E[u_{it}y_{i1}] & = 0, & t = 2, \ldots, T. \end{array} \tag{2}$$

Under these assumptions, the moments of the $T(T-1)$ interactions of Δy_{it} and y_{it} :

$$E[\Delta y_{it} y_{ij}], \qquad j = 1, \ldots, T, t = 2, \ldots, T \tag{3}$$

can be used to construct functions which identify the parameter of interest θ. We do not use products of Δy_{it} to identify θ, since we would need further assumptions, i.e., homoskedasticity or initial condition assumptions (see, e.g., Han and Phillips, 2010).

Two different sets of moment conditions, which are functions of the moments in (3), are commonly used to identify θ :

1. Dif moment conditions:

$$E[y_{ij}(\Delta y_{it} - \theta \Delta y_{it-1})] = 0, \qquad j = 1, \ldots, t-2; t = 3, \ldots, T, \tag{4}$$

as proposed by, e.g., Anderson and Hsiao (1981) and Arellano and Bond (1991). The Dif moment conditions solely result from the conditions in (2).

2. Lev moment conditions:

$$E[\Delta y_{it-1}(y_{it} - \theta y_{it-1})] = 0, \qquad t = 3, \ldots, T, \tag{5}$$

as proposed by Arellano and Bover (1995; see also Blundell and Bond, 1998). In addition to the conditions in (2), the Lev moment conditions use

$$E[\Delta y_{it} c_i] = 0, \tag{6}$$

which implies that the original data in levels have constant correlation over time with the individual-specific effects. The Lev moment conditions (5) hold under

[2]The extension to other explanatory variables would depend on the nature of these. For some settings, such an extension would be trivial, but for others not so.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

the following conditions regarding the initial observations y_{i1} $(i = 1, \ldots, N)$:

$$y_{i1} = \mu_i + u_{i1}, \tag{7}$$

$$\mu_i = c_i/(1-\theta), \tag{8}$$

$$\begin{aligned} E[u_{i1}] &= 0, \\ E[u_{i1}c_i] &= 0, \\ E[u_{i1}u_{it}] &= 0, \quad t > 1. \end{aligned} \tag{9}$$

The specification of the initial observations in (7)–(9) is often referred to as mean stationarity. In our analysis, we maintain the assumption of mean stationarity.

The Dif and Lev moments can be used separately or jointly to identify θ. When we use the moment conditions in (4) and (5) jointly, we refer to them as Sys moment conditions[3] (see Arellano and Bover, 1995; Blundell and Bond, 1998). Another set of NL moment conditions, which just like the Dif moments only use the conditions in (2), results from Ahn and Schmidt (1995):

$$E[(y_{it} - \theta y_{it-1})(\Delta y_{it-1} - \theta \Delta y_{it-2})] = 0, \qquad t = 4, \ldots, T. \tag{10}$$

The NL moments can be used separately or jointly with the Dif moments to identify θ. When we use the moment conditions in (4) and (10) jointly, we refer to them as AS moment conditions.

Ahn and Schmidt (1995) show that their AS moment conditions exhaust the information on θ in the moment conditions (2) and are therefore complete. Mean stationarity adds one moment condition (6) to the moment conditions in (2). Hence, the complete set of moment conditions under (2) and (6) equals the AS moment conditions and (6). Upon rewriting, we can show that these combined moment conditions are identical to the Sys moment conditions, so they are complete under (2) and (6).

2.2. Identification Robust GMM Tests

In GMM, we consider a k-dimensional vector of moment conditions (see Hansen, 1982):

$$E[f_i(\theta_0)] = 0, \qquad i = 1, \ldots, N, \tag{11}$$

where $f_i(\theta)$ is a k-dimensional (continuous and continuously differentiable) function of the observed data for individual i and the unknown parameter vector θ whose functional expression is identical for all individuals. There is a unique true value of the p-dimensional vector θ where the moment conditions are satisfied,

[3] We could extend the Lev moment conditions to $\frac{1}{2}(T-1)(T-2)$ sample moments by including additional interactions of Δy_{it-j} and $y_{it} - \theta y_{it-1}$, for $j = 2, \ldots, t-2$. It can be shown, however, that all conditions on top of those in (5) can be constructed as linear combinations of the Dif conditions in (4) and the Lev conditions in (5).

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

which we denote by θ_0, and k is at least as large as p. We only analyze the first-order autoregressive panel data model, so $p = 1$ for our setting. The population moments in (11) are estimated using the sample moments,

$$f_N(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} f_i(\theta). \tag{12}$$

The $k \times p$ dimensional matrix $q_N(\theta)$ contains the derivative of $f_N(\theta)$ with respect to θ :

$$q_N(\theta) = \tfrac{\partial}{\partial \theta'} f_N(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} q_i(\theta), \tag{13}$$

with $q_i(\theta) = \frac{\partial}{\partial \theta'} f_i(\theta)$. Specifications of the sample moment functions $f_N(\theta)$ and $q_N(\theta)$ for the Dif, Lev, Sys, NL, and AS moment conditions are provided in the Appendix.

Statistical inference based on the two-step GMM estimator is known to be of poor quality in the case of weak identification, which leads to an inconsistent estimator with nonstandard behavior of its corresponding t-statistic (see, e.g., Phillips, 1989; Staiger and Stock, 1997; Stock and Wright, 2000). The nonstandard limiting behavior of one-step and two-step GMM estimators for dynamic panel data models in the case of weak identification has been documented in, e.g., Madsen (2003), Kruiniger (2009), and Phillips (2018).

In this study, we therefore use identification robust GMM statistics to overcome the aforementioned problems. The main advantage of identification robust statistics is that, unlike conventional two-step GMM statistics, their limiting distributions are unaffected by the identification strength. Define θ^* as the hypothesized value under the null hypothesis. A particularly simple to compute identification robust GMM statistic to test $H_0 : \theta = \theta^*$ is the GMM extension of the AR statistic (see Anderson and Rubin, 1949; Stock and Wright, 2000):

$$GMM\text{-}AR(\theta^*) = N f_N(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} f_N(\theta^*), \tag{14}$$

with $\hat{V}_{ff}(\theta)$ the Eicker–White covariance matrix estimator:

$$\hat{V}_{ff}(\theta) = \tfrac{1}{N}\sum_{i=1}^{N} (f_i(\theta) - f_N(\theta))(f_i(\theta) - f_N(\theta))'. \tag{15}$$

The GMM-AR statistic equals the continuous updating objective function (Hansen, Heaton, and Yaron, 1996) evaluated in θ^*. A possible drawback of the GMM-AR statistic is its lower power in the case of overidentified models. The KLM statistic of Kleibergen (2005) partly overcomes this. The KLM statistic is a quadratic form of the score of the GMM-AR statistic with respect to θ:

$$\begin{aligned} KLM(\theta^*) = {} & N f_N(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} \hat{D}_N(\theta^*) \left[\hat{D}_N(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} \hat{D}_N(\theta^*) \right]^{-1} \\ & \hat{D}_N(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} f_N(\theta^*), \end{aligned} \tag{16}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

with $\hat{D}_N(\theta)$ a $k \times p$ dimensional matrix,

$$\text{vec}(\hat{D}_N(\theta)) = \text{vec}(q_N(\theta)) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_N(\theta), \tag{17}$$

and

$$\hat{V}_{qf}(\theta) = \tfrac{1}{N}\sum_{i=1}^{N} (\text{vec}[q_i(\theta) - q_N(\theta)])\,(f_i(\theta) - f_N(\theta))'. \tag{18}$$

The limiting distributions of the identification robust GMM-AR and KLM statistics apply under less restrictive assumptions than those of the traditional test statistics based on two-step GMM. The GMM-KLM and GMM-AR statistics converge under H_0 to $\chi^2(p)$ and $\chi^2(k)$ distributed random variables even when the Jacobian, $J(\theta_0) = E(q_i(\theta_0))$, does not have a full-rank value (see Stock and Wright, 2000; Kleibergen, 2005; Newey and Windmeijer, 2009). Other identification robust statistics for GMM are proposed in Kleibergen (2005), Andrews (2016), and Andrews and Mikusheva (2016), which all provide extensions of the conditional likelihood ratio statistic of Moreira (2003) to GMM. The conditional likelihood ratio statistic is optimal for the homoskedastic linear instrumental variables regression model with one included endogenous variable (see Andrews, Moreira, and Stock, 2006). None of its extensions to GMM has, however, shown to be optimal for our setting of the dynamic linear panel autoregression, so we just use the easier to implement GMM-AR and KLM statistics.[4]

The identification robust tests can be inverted to obtain corresponding identification robust confidence sets. The $100 \times (1-\alpha)\%$ confidence set for θ (denoted by $\text{CS}_\theta(\alpha)$ below) consists of all values of θ^* for which the respective identification robust test does not reject using its $100 \times \alpha\%$ asymptotic critical value:

$$\text{CS}_\theta(\alpha) = \{\theta^* : IRT(\theta^*) \leq CDF_{IRT}(\alpha)\}, \tag{19}$$

with $IRT(\theta^*)$ the identification robust statistic evaluated at θ^* and $CDF_{IRT}(\alpha)$ the $(1-\alpha) \times 100$th percentile of the limiting distribution of $IRT(\theta_0)$.

The identification robust tests are not quadratic functions of θ^*, so they cannot directly be inverted to obtain the confidence set.[5] The confidence sets resulting from them do, therefore, not have the usual expression of an estimator plus or minus a multiple of the standard error. Instead, we have to specify a p-dimensional grid of values of θ^* and compute the identification robust statistic for every value of θ^* on the grid to determine if it is less than the appropriate critical value, so θ^* is part of the confidence set.

Specifically, the confidence set in (19) can have three distinct shapes:

[4] Andrews et al. (2006) establish the optimality of the likelihood ratio test for the i.i.d. linear instrumental variable regression model using the Neymann–Pearson lemma. We cannot do so here, since the identification of θ depends on other nuisance parameters besides the Jacobian, like the initial observations, so it is not obvious how optimality can be established.

[5] An exception is the GMM-AR statistic in the homoskedastic linear instrumental variable regression model (see Dufour and Taamouti, 2005).

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

1. Bounded and convex: there is a closed compact set of values of θ^* for which the identification robust test statistic does not exceed the critical value.
2. Unbounded: this occurs either when there are no values of θ^* for which the identification robust test statistic exceeds the critical value (unbounded and convex), or when there are bounded sets of values of θ^* for which the identification robust test statistic exceeds the critical value (unbounded and disjoint).
3. Empty: the identification robust test statistic exceeds the critical value for all values of θ^*.

Bounded and convex confidence sets occur when the parameters of interest are well identified. Unbounded confidence sets are indicative of weak identification, so if we then test $H_0 : \theta = \theta^*$ at a very large, possibly infinite, value of θ^* using an identification robust test at, say, the 5% significance level, it does not necessarily reject. For such instances, we thus often do not reject the hypothesis of an infinite value of θ, so we obtain an unbounded 95% confidence set. In Dufour (1997, Thms. 3.3 and 3.6), it is shown that any size correct procedure used to test parameters which can be nonidentified must have a positive probability of producing an unbounded 95% confidence set. Conversely, also any test procedure, like the Wald t test, which cannot generate an unbounded 95% confidence set, cannot be a size correct test procedure when the tested parameter can be nonidentified. Empty confidence sets occur when the model is misspecified, so there is no value of θ for which the moment condition holds. Since the GMM-AR statistic tests whether all moment conditions hold, it also tests misspecification. It can therefore result in empty confidence sets, but the KLM test cannot, since it is equal to zero at the continuous updating estimator of Hansen et al. (1996), which is the minimizer of the GMM-AR statistic.

The identification robust statistics conduct tests on the full parameter vector θ. Valid $(1-\alpha) \times 100\%$ confidence sets for the individual elements of θ then result by projecting the joint p-dimensional $(1-\alpha) \times 100\%$ confidence set for θ on the p different axes. These projection-based confidence sets are size correct, so they contain the true value of θ with a probability which is at least $(1-\alpha) \times 100\%$ irrespective of the strength of identification. Projection-based confidence sets can face computational issues when p is rather large given the large number of points on the p-dimensional grid for which the statistic then has to be computed.

Confidence sets for the individual elements of θ can also be obtained by plugging in an estimator for the remaining elements of θ after which the (conditional) limiting distribution can be sharpened using the usual degrees of freedom correction of the χ^2 limiting distributions. The resulting confidence sets only have correct coverage when these remaining parameters are well identified (see Kleibergen, 2005). Just in some isolated cases, for example, when using the GMM-AR statistic in the homoskedastic linear instrumental variables regression model or in the linear factor model for determining risk premia in finance, can we prove that these confidence sets are valid without requiring the partialled out parameters

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

to be well identified (see Guggenberger et al., 2012; Guggenberger, Kleibergen, and Mavroeidis, 2019; Kleibergen (2021); Kleibergen, Kong, and Zhan (2020); Kleibergen and Zhan (2020)).

2.3. Using Identification Robust Tests to Highlight Identification Issues

Identification robust GMM tests are size correct irrespective of the identification strength. Therefore, their rejection frequencies can be used in a straightforward manner to illustrate the identification issues at particular values of the autoregressive parameter in the dynamic panel data model. The conventional t test based on the two-step GMM estimator is not suitable for this purpose, as it is size distorted in the case of weak identification and, hence, rejection frequencies would not equal the significance level.

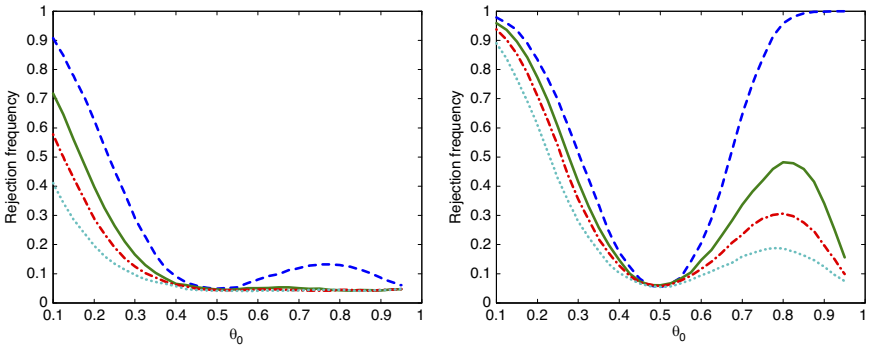
To illustrate the identification issues for the different moment conditions, we compute the rejection frequencies of 5% significance KLM tests of $H_0 : \theta = 0.5$ for a range of (true data generating) values θ_0. We do so by simulating data from the panel autoregressive model in (1) with three or four time series observations, so $T = 3$ or 4, and 250 individuals, so $N = 250$. The individual specific effects c_i and idiosyncratic errors u_{it} are independently generated from $N(0, \sigma_c^2)$ and $N(0, 1)$ distributions, respectively. We vary the value of σ_c^2 to show the sensitivity of the identification of θ using the panel moment conditions to the variance of the initial observations. We assume mean stationarity, so (7)–(9) hold.

We consider four KLM tests based on Dif, Lev, Sys, and AS moment conditions, which have been calculated according to equation (16) using $\theta^* = 0.5$. Figures 1 and 2 show the rejection frequencies of KLM tests of $H_0 : \theta = 0.5$ with 5% significance for four values of σ_c^2 and a range of true values θ_0. Figure 1 does so for three times series observations, while Figure 2 covers four time series observations. The simulation experiment is designed such that the variance of the initial observations becomes very large when θ_0 gets close to one and σ_c^2 exceeds zero.

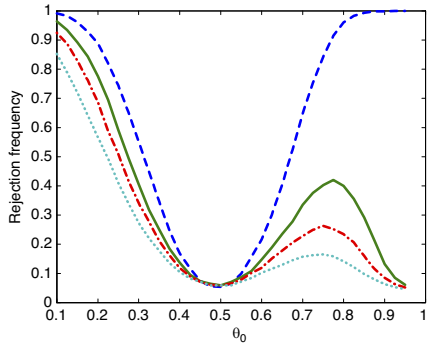
Figures 1a and 2a show that the rejection frequencies of the KLM test with Dif moment conditions for θ_0 close to one converges to the significance level of 5%. It is well known that the Jacobian of the Dif moment conditions is zero when θ_0 equals one, so they then do not identify θ. The KLM test is identification robust, which explains why the rejection frequency equals the significance level both at the hypothesized value of $\theta^* = 0.5$ and when θ_0 is close to 1 for all values of σ_c^2. The latter results, since the Dif moment conditions do then not identify θ; hence, the KLM test has no discriminating power, so the power of the KLM test equals the significance level.

Figures 1b and 2b show the rejection frequencies of 5% significance tests of $H_0 : \theta = 0.5$ using the KLM test with Lev moment conditions. Interestingly, these figures show that the Lev moment conditions only identify θ when the true value θ_0 is close to one when $\sigma_c^2 = 0$. Nonzero values of σ_c^2 correspond with a large variance of the initial observations when θ_0 is close to one and Figures 1b and

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press



(a) Dif moment conditions. (b) Lev moment conditions.



(c) Sys moment conditions.

FIGURE 1. Rejection frequencies of KLM test of $H_0 : \theta = 0.5$ with 5% significance using different moment conditions for $T = 3, N = 250$, and $\sigma_c^2 = 0$ (dashed), 0.5 (solid), 1 (dash-dotted), and 2 (dotted).

2b show that the Lev moment conditions do not identify θ in this case. This contradicts the common perception that the Lev moment conditions generally identify θ irrespective of the setting of nuisance parameters, like, the variance of the initial observations.

Figures 1c and 2c show the rejection frequencies of 5% significance tests of $H_0 : \theta = 0.5$ using the KLM test with Sys moment conditions. Surprisingly, these figures show that the Sys moment conditions do not identify θ when θ_0 is close to one and $\sigma_c^2 > 0$ when $T = 3$, but do so when $T = 4$.

Figure 2d shows the rejection frequencies of 5% significance tests of $H_0 : \theta = 0.5$ using the KLM test with AS moment conditions. These rejection frequencies show that the AS moment conditions, which are not defined for $T = 3$, identify θ when its true value is close to one and the variance of the initial observations is very large. Interestingly, the rejection frequencies of KLM tests of H_0 using the Sys and AS moment conditions are very close when θ_0 is near one when paired with large variances of the initial observations.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

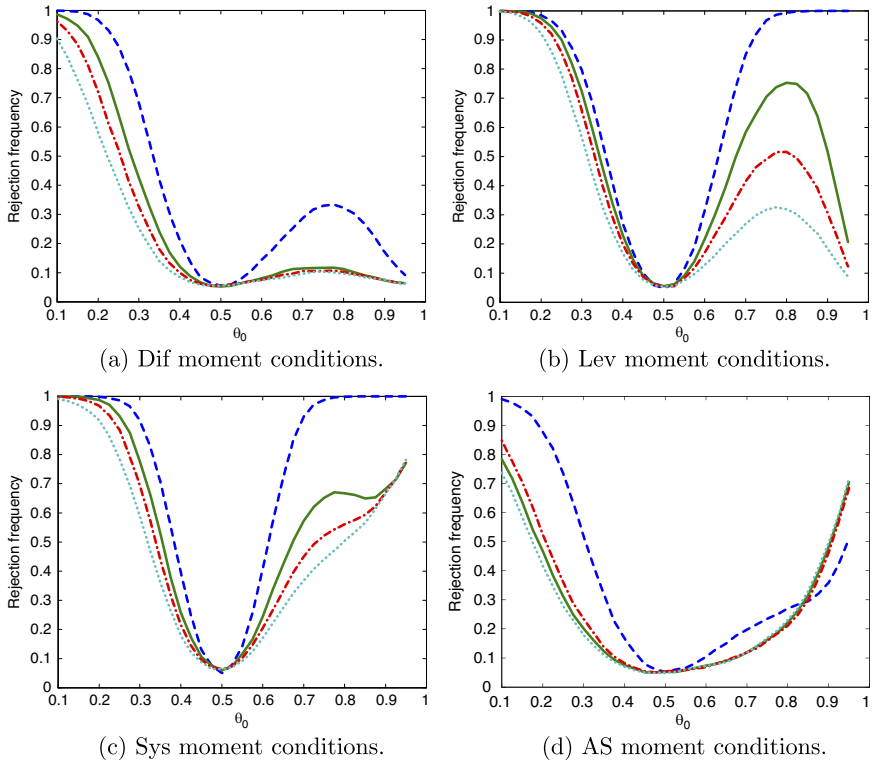


FIGURE 2. Rejection frequencies of KLM test of $H_0 : \theta = 0.5$ with 5% significance using different moment conditions for $T = 4, N = 250$, and $\sigma_c^2 = 0$ (dashed), 0.5 (solid), 1 (dash-dotted), and 2 (dotted).

Summarizing, Figures 1 and 2 illustrate a few stylized facts that concern the identification of θ for the data generating process (DGP) used in the simulation experiment:

1. Dif moment conditions do not identify θ when θ_0 is close to one for general T.
2. Lev moment conditions do not identify θ when θ_0 is close to one for large variances of the initial observations for general T.
3. Sys moment conditions do not identify θ when θ_0 is close to one for large variances of the initial observations when $T = 3$.
4. Sys and AS moment conditions identify θ when θ_0 is close to one for large variances of the initial observations when T exceeds 3.
5. The rejection frequencies of KLM tests of H_0 using AS and Sys moment conditions when θ_0 is close to one and the variance of the initial observations is large are almost identical.

Except for the first stylized fact, a theory backing them up is lacking, so we aim to provide one in the sections ahead. In doing so, we show that all

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

information regarding θ, when its true value is close to one and the variance of the initial observations is large, is contained in a set of, so-called, robust moment conditions which are a combination of either the AS or Sys moment conditions. We furthermore show that the KLM test based on the original AS or Sys moment conditions, as reported in Figures 1 and 2, makes optimal use of these robust sample moments when only they contain information on θ.

Alongside the identification issues we can infer from the rejection frequencies in Figures 1 and 2, they are also indicative of the different kind of confidence sets that can result from the identification robust tests as discussed previously. For example, the low rejection frequencies occurring for θ_0 around one, that result from the identification issues, show that the 95% confidence sets for θ are then typically very wide, possibly unbounded, when θ_0 has such a value paired with a large variance of the initial observations. To visualize this further, Figure 3 contains the (one minus the) p-value plots of KLM tests using AS, Dif, Lev, and Sys moment

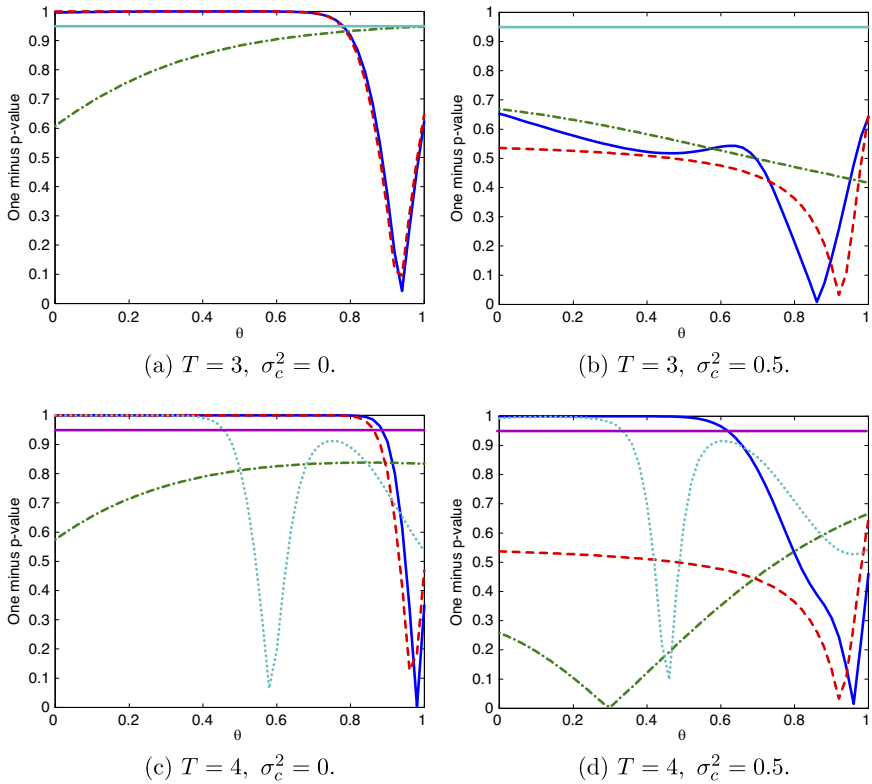


FIGURE 3. One minus p-value plots of KLM tests using different moments conditions: Sys (solid), AS (dotted), Lev (dashed), and Dif (dash-dot) for $\theta_0 = 0.95$ and $N = 250$.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

conditions for four datasets using the same DGPs as in Figures 1 and 2 with $N = 250$ and $\theta_0 = 0.95$.[6] The DGPs used for the four figures differ over the values of T and σ_c^2. The intersections of the depicted p-value plots with the line at 0.95 indicate the 95% confidence sets of KLM tests with the respective moment condition.

In Figure 3a,c, $\sigma_c^2 = 0$, so identification issues only occur at θ_0 close to one when using the Dif moment conditions. Since θ_0 is 0.95, this explains why the p-value plots of the KLM test with the Dif moments conditions do not cross the line at 0.95 in Figure 3a,c, so the resulting 95% confidence sets are very wide. The p-value plots in Figure 3a,c of KLM tests with Sys and Lev moment conditions show that they lead to bounded 95% confidence sets, since these moment conditions have no identification issues when $T = 3$ and $\sigma_c^2 = 0$.

In Figure 3b, where $T = 3$ and $\sigma_c^2 = 0.5$, none of the p-value plots crosses the line at 0.95, so 95% confidence sets that result from KLM tests with Dif, Lev, and Sys moment conditions are all very wide and possibly unbounded. This is indicative of the identification issues when $T = 3$ and $\sigma_c^2 = 0.5$ for true values of θ close to one.

In Figure 3d, where $T = 4$ and $\sigma_c^2 = 0.5$, KLM tests with Sys and AS moment conditions both result in finite 95% confidence sets, while the KLM test with Dif and Lev moment conditions leads to very wide possibly unbounded confidence sets. Hence, Sys and AS moment conditions have no identification issues, while Dif and Lev moment conditions do. The AS moment conditions are quadratic functions of θ, which explains the somewhat unusual shape of their p-value plots in Figure 3c,d.

3. IDENTIFICATION FROM DIFFERENT MOMENT CONDITIONS

Stylized Facts 1–4 illustrated by Figures 1–3 show the identification issues that occur for the autoregressive parameter θ when the variance of the initial observations is large and θ_0, i.e., the true value in the DGP, is close to one. To pin these identification issues down precisely, we use an asymptotic sampling scheme which consists of joint drifting sequences for the autoregressive parameter and the variance of the initial observation. We indicate this dependence on the sample size N by $\theta_{0,N}$ and $h_N(\theta_{0,N}) = \frac{1}{\sqrt{var(y_{i1})}}$, respectively. The true value of θ, previously denoted by θ_0, is from now on, therefore, denoted by $\theta_{0,N}$. Assumptions 1 and 2 group the different requirements needed to obtain our results.

Assumption 1. a. The drifting sequences of the autoregressive parameter and variance of the initial observations are such that:

$$\begin{aligned} \lim_{N\to\infty} \theta_{0,N} &= 1, \\ \lim_{N\to\infty} h_N(\theta_{0,N}) &= d_1, \end{aligned} \tag{20}$$

[6] We note that the Figure 3a–d shows (one minus) the p-value for one realized dataset and does not show the simulated empirical distribution function of the test under the null hypothesis, which is sometimes also referred to as a p-value plot (see Davidson and MacKinnon, 2002).

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

with d_1 a finite, possibly zero constant.

b. The initial observations satisfy the mean stationarity conditions in (7)–(9).

c. The joint limit behavior of the variance of u_{i1} and $(1-\theta_{0,N})$ is such that

$$\lim_{N\to\infty}(1-\theta_{0,N})\sigma^2_{1,N} = d_2, \tag{21}$$

with $\sigma^2_{1,N} = \text{var}(u_{i1})$, d_2 a finite, possibly zero constant, and $(1-\theta_{0,N})^{\frac{1}{2}}u_{i1}$ is a random variable with finite fourth-order moments.

d. The variance of the product of the initial observation y_{i1} and the disturbances u_{it} is such that

$$var(u_{it}y_{i1}) = \sigma_t^2 var(y_{i1}), \quad t = 2,\ldots,T, \tag{22}$$

with $\sigma_t^2 = var(u_{it}), \quad t = 2,\ldots,T$.

e. The errors $u_{i1}/\sigma_{1,N}$, $u_{i2},\ldots,u_{iT}$ and c_i, $i = 1,\ldots,N$, are independently distributed within individuals and over the different individuals and have mean-zero, finite variance, and finite fourth-order moments and satisfy the conditions in (2).

Assumption 1(a) concerns the joint limit behavior of the variance of the initial observations and $\theta_{0,N}$. By the definition of μ_i in (8) and Assumption 1(a), μ_i is also drifting with the sample size, since it is a function of $\theta_{0,N}$, and so are y_{i1} and $\sigma^2_{1,N}$. Assumption 1(b) specifies that the initial observations follow the mean stationarity assumption, which is necessary for the Lev and Sys moment conditions to hold. Assumption 1(c)–(e) is mainly technical assumptions, which is needed to obtain our theoretical results. Assumption 1(c) sets an upper bound on the rate at which the variance of u_{i1} can diverge. It implies that the variance of u_{i1} is at most proportional to $(1-\theta_{0,N})^{-1}$ (so covariance stationarity is allowed for). Assumption 1(d) holds under independence of u_{it} and y_{i1}, but it can also hold under less stringent conditions. In the sequel, we analyze the identification of θ when the variance of the initial observations gets large compared to that of the subsequent disturbances. Assumption 1(d) enables such settings. Assumption 1(e) is a technical assumption, which is needed to use a central limit theorem.

Assumption 1(a) allows the variance of the initial observations to be large jointly with a large value for the autoregressive parameter. When d_1 in (20) equals zero, the rate at which $h_N(\theta_{0,N})$ goes to zero, or the variance of the initial observation goes to infinity, is key to the identification of θ from the sample moment conditions. We therefore put down two alternative assumptions regarding the joint convergence of the sample size and the variance of the initial observations under which there is identification or identification is problematic for specific moment conditions.

Assumption 2. a. $d_1 = 0$ and the drifting sequence of the variance of the initial observation is such that:

$$h_N(\theta_{0,N})\sqrt{N} \underset{N\to\infty}{\to} 0. \tag{23}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

b. $d_1 \neq 0$ or the drifting sequence of the variance of the initial observation is such that:

$$h_N(\theta_{0,N})\sqrt{N} \underset{N\to\infty}{\to} \infty. \tag{24}$$

Identification generically holds under Assumption 2(b) but can become problematic under Assumption 2(a) and then depends on the particular moment condition and number of time series observations as we show later on. In the intermediate case where $h_N(\theta_{0,N})\sqrt{N}$ converges to a finite, but nonzero constant, we are in a case similar to that discussed in the weak instrument literature where the sample Jacobian converges to a random variable which leads to inconsistent estimators with nonstandard behavior of their corresponding t-statistics. Because of the practical similarities with Assumption 2(a), however, we do not separately discuss it.

Since any assumption about the convergence rates of the sample size and the variance of the initial observations is to a large extent arbitrary, also the identification of θ by these conditions is arbitrary for DGPs for which the true value of θ is close to one and the variance of the initial observations is infinite when the true value of θ equals one. Some plausible DGPs, all of which accord with mean stationarity (7)–(9), for the initial observations belong to this category:

DGP 1. $\sigma_c^2 = \text{var}(c_i)$, $\sigma_{1,N}^2 = \sigma_1^2$, $h(\theta_{0,N})^{-2} = \sigma_c^2/(1-\theta_{0,N})^2 + \sigma_1^2$, so when $\theta_{0,N} \underset{N\to\infty}{\to} 1$, $(1-\theta_{0,N})^{-1}h(\theta_{0,N}) \underset{N\to\infty}{\to} \sigma_c^{-1}$.

DGP 2. $\sigma_c^2 = \text{var}(c_i)$, $\sigma_{1,N}^2 = \frac{\sigma^2}{1-\theta_{0,N}^2}$, $\sigma^2 = var(u_{it})$, $t = 2,\ldots,T$, $h(\theta_{0,N})^{-2} = \sigma_c^2/(1-\theta_{0,N})^2 + \sigma^2/(1-\theta_{0,N}^2)$, so when $\theta_{0,N} \underset{N\to\infty}{\to} 1$, $(1-\theta_{0,N})^{-1}h(\theta_{0,N}) \underset{N\to\infty}{\to} \sigma_c^{-1}$.

DGP 3. $\sigma_\mu^2 = \text{var}(\mu_i)$, $\sigma_{1,N}^2 = \frac{\sigma^2}{1-\theta_{0,N}^2}$, $\sigma^2 = var(u_{it})$, $t = 2,\ldots,T$, $h(\theta_{0,N})^{-2} = \sigma_\mu^2 + \sigma^2/(1-\theta_{0,N}^2)$, so when $\theta_{0,N} \underset{N\to\infty}{\to} 1$, $(1-\theta_{0,N}^2)^{-\frac{1}{2}}h(\theta_{0,N}) \underset{N\to\infty}{\to} \sigma^{-1}$.

DGP 4. $\sigma_\mu^2 = \text{var}(\mu_i)$, $\sigma_{1,N}^2 = \sigma^2\frac{1-\theta_{0,N}^{2(g+1)}}{1-\theta_{0,N}^2}$, $\sigma^2 = var(u_{it})$, $t = 2,\ldots,T$, $h(\theta_{0,N})^{-2} = \sigma_\mu^2 + \sigma^2\frac{1-\theta_{0,N}^{2(g+1)}}{1-\theta_{0,N}^2}$, so when $\theta_{0,N} \underset{N\to\infty}{\to} 1$, $\left(\frac{1-\theta_{0,N}^2}{1-\theta_{0,N}^{2(g+1)}}\right)^{-\frac{1}{2}} h(\theta_{0,N}) \underset{N\to\infty}{\to} \sigma^{-1}$.

DGP 5. $\sigma_c^2 = \text{var}(c_i)$, $\sigma_{1,N}^2 = \sigma^2\frac{1-\theta_{0,N}^{2(g+1)}}{1-\theta_{0,N}^2}$, $\sigma^2 = var(u_{it})$, $t = 2,\ldots,T$, $h(\theta_{0,N})^{-2} = \sigma_c^2/(1-\theta_{0,N})^2 + \sigma^2\frac{1-\theta_{0,N}^{2(g+1)}}{1-\theta_{0,N}^2}$, so when $\theta_{0,N} \underset{N\to\infty}{\to} 1$, $\left(1-\theta_{0,N}\right)^{-1} h(\theta_{0,N}) \underset{N\to\infty}{\to} \sigma_c^{-1}$.

DGPs 4 and 5 characterize an autoregressive process of order one that has started g periods in the past, while the initial observations that result from DGP 2 and 3

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

result from an autoregressive process that has started an infinite number of periods in the past. DGPs 2 and 3 are also used by Blundell and Bond (1998), and Arellano and Bover (1995) use DGP 2, but these studies keep the variance of the initial observations fixed.

For DGPs 1–5 to imply Assumption 2(a), the limiting sequence $\theta_{0,N}$ has to be such that:

$$\begin{aligned}
&\text{DGP } 1,2,5: \quad (1-\theta_{0,N})\sqrt{N} \underset{N\to\infty}{\to} 0 \quad \text{for which it is sufficient that } \theta_{0,N} = 1 - \tfrac{e}{N^{\frac{1}{2}(1+\epsilon)}},\\
&\text{DGP } 3: \quad (1-\theta_{0,N}^2)N \underset{N\to\infty}{\to} 0 \quad \text{for which it is sufficient that } \theta_{0,N} = 1 - \tfrac{e}{N^{1+\epsilon}},\\
&\text{DGP } 4: \quad \tfrac{N}{g} \underset{N\to\infty, g\to\infty}{\to} 0,
\end{aligned} \tag{25}$$

with e a constant and ϵ some real number larger than zero. In the case of DGP 4, (25) implies that the process has been running longer than the sample size N. Kruiniger (2009) uses the above specification of DGP 3 with $\epsilon = 0$ and DGP 4 with N/g converging to a constant to construct local to unity asymptotic approximations of the distributions of two-step GMM estimators that use the Dif, Lev, or Sys moment conditions.

We do not confine ourselves to a specific DGP for the initial observations, so we obtain results that apply more generally. While the (non) identification conditions for identifying θ that result from the above DGPs might be (in)plausible, it is the arbitrariness of them which is problematic. In addition, the identification condition might hold, but it can still lead to large size distortions of Wald test statistics, like, the t test.

To analyze the identification of θ by the different moment conditions for a general number of time periods T, we start out with a representation theorem. For the different moment conditions, it states the behavior of the sample moments and their derivatives under Assumptions 1 and 2(a).

THEOREM 1 (Representation theorem). *Under Assumptions 1 and 2(a), we can characterize the large sample behavior of the Dif, Lev, NL, AS, and Sys sample moments for T time series observations and their derivatives by:*

$$\begin{aligned}
\begin{pmatrix} f_N^j(\theta) \\ q_N^j(\theta) \end{pmatrix} = \begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix} \left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}} \left(\psi - h_N(\theta_{0,N})\sigma_{1,N}\iota_{T-1}\psi_c\right) - \iota_{T-1}d_2 \right] \\
+ \begin{pmatrix} \mu_f^j(\theta, \bar{\sigma}^2) \\ \mu_q^j(\theta, \bar{\sigma}^2) \end{pmatrix} + o_p(1),
\end{aligned} \tag{26}$$

with j = Dif, Lev, NL, AS, Sys. The specifications of the k_j-dimensional sample moments $f_N^j(\theta)$ and derivatives $q_N^j(\theta)$ are given in the Appendix. Furthermore, $A_f^j(\theta)$, $A_q^j(\theta)$, $\mu_f^j(\theta, \bar{\sigma}^2)$, and $\mu_q^j(\theta, \bar{\sigma}^2)$ are constant $k_j \times (T-1)$, $k_j \times (T-1)$, $k_j \times 1$,

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

and $k_j \times 1$ *dimensional matrices,* $\bar{\sigma}^2 = (\sigma_2^2 \dots \sigma_T^2)$,

$$\begin{aligned} \frac{h_N(\theta_{0,N})}{\sqrt{N}} \sum_{i=1}^{N} \begin{pmatrix} y_{i1}u_{i2} \\ \vdots \\ y_{i1}u_{iT} \end{pmatrix} &\xrightarrow[d]{} \quad \psi, \\ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{u_{i1}}{\sigma_{1,N}} c_i &\xrightarrow[d]{} \quad \psi_c, \end{aligned} \tag{27}$$

so ψ *is a* $(T-1)$*-dimensional normal random vector,* $\psi \sim N(0, diag(\sigma_2^2 \dots \sigma_T^2))$, $\psi_c \sim N(0, var(c_i))$ *and independent from* ψ*, and* ι_{T-1} *is a* $(T-1)$*-dimensional vector of ones. The specifications of* $A_f^j(\theta)$*,* $A_q^j(\theta), \mu_f^j(\theta, \bar{\sigma}^2)$*, and* $\mu_q^j(\theta, \bar{\sigma}^2)$ *for values of T equal to 3–5 are all stated in the Appendix.*

Proof. See the Appendix. □

The representation theorem in Theorem 1 is reminiscent of the cointegration representation theorem (see, e.g., Engle and Granger, 1987 and Johansen, 1991). Identical to that representation theorem, Theorem 1 shows that the behavior of the moment series changes over different directions.

Theorem 1 implies that the sample moment and its derivative diverge in the direction of $\binom{A_f^j(\theta)}{A_q^j(\theta)}$, since the latter components get multiplied by $\frac{1}{h(\theta_{0,N})\sqrt{N}}$, which under Assumption 2(a) goes off to infinity when the sample size increases. The only identifying information for θ then results from that part of the sample moment which does not depend on ψ. Since ψ only affects the part of the sample moments spanned by $A_f^j(\theta)$, the sample moments are independent of ψ in the direction of the maximal nondegenerate space spanned by vectors orthogonal to $A_f^j(\theta)$ to which we refer as the orthogonal complement of $A_f^j(\theta)$. We construct the orthogonal complement, which we denote by $A_f^j(\theta)_\perp$, as the full-rank matrix projecting on the orthogonal complement of the range space of $A_f^j(\theta)$. It consists of the minimal set of vectors spanning the null space of the columns of $A_f^j(\theta)$. In the case the null space has dimension zero, a full-rank specification of $A_f^j(\theta)_\perp$ cannot be constructed.

When we premultiply the sample moments by the orthogonal complement of $A_f^j(\theta)$, we obtain

$$A_f^j(\theta)'_\perp f_N^j(\theta) = \quad A_f^j(\theta)'_\perp \mu_f^j(\theta, \bar{\sigma}^2) + o_p(1). \tag{28}$$

Compared with expression (26) in Theorem 1, the elements multiplied by $A_f^j(\theta)$ have dropped out, since $A_f^j(\theta)'_\perp A_f^j(\theta) \equiv 0$. The right-hand side of (28) now contains all the remaining identifying elements of the original moment conditions. From expression (28), it is seen that identification results only when (1) $A_f^j(\theta)_\perp$ is a full-rank matrix; and (2) $A_f^j(\theta)'_\perp \mu_f^j(\theta, \bar{\sigma}^2) \neq 0$, for all $\theta \neq \theta_{0,N}$.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

For an illustrative example of Theorem 1, consider the large sample behavior, for $T = 3$ of the Lev sample moment, $\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}(y_{i3} - \theta y_{i2})$, and its derivative, $-\frac{1}{N}\sum_{i=1}^{N} y_{i2}\Delta y_{i2}$, when $\theta_{0,N}$ converges to one according to (20) and mean stationarity (8)–(9) applies. The Lev moment condition has been proposed by Arellano and Bover (1995) and Blundell and Bond (1998) to overcome the identification problems of the Dif moment condition near the unit root. Under Assumption 1, the relevant elements for the large sample behavior are:

$$\begin{aligned} f_N^{Lev}(\theta) &= \tfrac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}(y_{i3} - \theta y_{i2}) \\ &= (1-\theta)\left\{\tfrac{1}{N}\sum_{i=1}^{N} u_{i2}^2 + \tfrac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1} + \tfrac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}y_{i1}\right\} + o_p(1), \\ q_N^{Lev}(\theta) &= -\tfrac{1}{N}\sum_{i=1}^{N} y_{i2}\Delta y_{i2} \\ &= -\tfrac{1}{N}\sum_{i=1}^{N} u_{i2}^2 - \tfrac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1} \\ &\quad -\tfrac{1}{N}\sum_{i=1}^{N}(1-\theta_{0,N})u_{i1}y_{i1} + o_p(1) \end{aligned} \tag{29}$$

(see the proof of Theorem 1 in the Appendix for a derivation). The $o_p(1)$ remainder terms contain all elements in (29) that cannot dominate the large sample behavior when $\theta_{0,N}$ goes to one according to the drifting parameter sequences defined in Assumption 1. The components explicitly specified in (29) either have a nonzero mean or depend on the initial observations y_{i1}. Under Assumption 1, we have that

$$h_N(\theta_{0,N})\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} u_{i2}y_{i1} \underset{d}{\to} \psi_2,\ \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\tfrac{u_{i1}}{\sigma_{1,N}}c_i \underset{d}{\to} \psi_c, \tag{30}$$

which is proved in Lemma 1 in the Appendix and where ψ_2 and ψ_c are independent normal random variables with mean zero and variance σ_2^2 and σ_c^2, $\sigma_c^2 = \mathrm{var}(c_i)$. It explains why $\frac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1}$ and $\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}y_{i1} = \frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}^2 + \frac{1}{N}\sum_{i=1}^{N} u_{i1}c_i$ explicitly appear in (29). When d_1 in (20) equals zero, the rate at which $h_N(\theta_{0,N})$ goes to zero, or the variance of the initial observation goes to infinity, determines the behavior of the sample moments in (29). For example, when $d_1 = 0$ and these sequences are as in Assumption 2(b), it holds that

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} y_{i2}\Delta y_{i2} \quad \underset{p}{\to} \quad \sigma_2^2 - d_2. \tag{31}$$

Although Assumption 1 does not fully pin down d_2, which value depends on the particular DGP for the initial observations, it is clear that the probability limit of the sample Jacobian typically differs from zero. Hence, the Lev moment condition seems to identify θ irrespective of its true value (see Arellano and Bover, 1995;

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Blundell and Bond, 1998). There is a caveat though, since, under Assumption 2(a), Theorem 1 shows that:

$$
\begin{aligned}
f_N^{Lev}(\theta) &= \frac{1}{h_N(\theta_{0,N})\sqrt{N}} \frac{h_N(\theta_{0,N})}{\sqrt{N}} \sum_{i=1}^{N} \Delta y_{i2}(y_{i3} - \theta y_{i2}) \\
&= (1-\theta)\left\{ \frac{1}{h_N(\theta_{0,N})\sqrt{N}} (\psi_2 - h_N(\theta_{0,N})\sigma_{1,N}\psi_c) + (\sigma_2^2 - d_2) \right\} + o_p(1), \\
q_N^{Lev}(\theta) &= -\frac{1}{h_N(\theta_{0,N})\sqrt{N}} \frac{h_N(\theta_{0,N})}{\sqrt{N}} \sum_{i=1}^{N} y_{i2}\Delta y_{i2}, \\
&= -\frac{1}{h_N(\theta_{0,N})\sqrt{N}} (\psi_2 - h_N(\theta_{0,N})\sigma_{1,N}\psi_c) - (\sigma_2^2 - d_2) + o_p(1), \qquad \textbf{(32)}
\end{aligned}
$$

which implies that the sample moments of the Lev population moment and Jacobian diverge when the sample size increases. The Lev sample moment then no longer identifies θ, since the components that would identify θ in the Jacobian identification condition, i.e., $\frac{1}{N}\sum_{i=1}^{N} u_{i2}^2$, gets dominated by the component $\frac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1}$ and possibly $\frac{1}{N}\sum_{i=1}^{N}(1-\theta_{0,N})u_{i1}y_{i1}$.

We next discuss what Theorem 1 implies for the different sets of moment conditions discussed previously and their respective orthogonal complements of $A_f(\theta)$.

Dif and Lev conditions.

When $T = 3$ or 4, the specifications of $\mu_f^j(\theta,\bar{\sigma}^2)$, $A_f^j(\theta)$, and $A_f^j(\theta)_\perp$ for the Dif and Lev moment conditions, which are stated in the proof of Theorem 1 in the Appendix, are:

$$
\begin{aligned}
\textbf{Dif:}\quad \mathbf{T=3}\quad & \mu_f^{Dif}(\theta,\bar{\sigma}^2) = 0, A_f^{Dif}(\theta) = (-\theta \;\; 1), A_f^{Dif}(\theta)_\perp = (1 \;\; \theta), \\
\mathbf{T=4}\quad & \mu_f^{Dif}(\theta,\bar{\sigma}^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix}, \\
& A_f^{Dif}(\theta)_\perp = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}. \\
\textbf{Lev:}\quad \mathbf{T=3}\quad & \mu_f^{Lev}(\theta,\bar{\sigma}^2) = (1-\theta)\begin{pmatrix} \sigma_2^2 \\ 0 \end{pmatrix}, A_f^{Lev}(\theta) = (1-\theta \;\; 0), A_f^{Lev}(\theta)_\perp \\
& \text{does not exist,}
\end{aligned}
$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$\mathbf{T=4} \quad \mu_f^{Lev}(\theta,\bar{\sigma}^2)=(1-\theta)\begin{pmatrix}\sigma_2^2\\ \sigma_3^2\\ 0\end{pmatrix}, A_f^{Lev}(\theta)=\begin{pmatrix}1-\theta & 0 & 0\\ 0 & 1-\theta & 0\end{pmatrix},$$
$$A_f^{Lev}(\theta)_\perp \text{ does not exist.} \tag{33}$$

The expressions of $A_f^{Lev}(\theta)$ are all such that we cannot specify a nonzero matrix $A_f^{Lev}(\theta)_\perp$ such that $A_f^{Lev}(\theta)'_\perp A_f^{Lev}(\theta)=0$. This remains so when T exceeds 4 (see the Appendix). Hence, $A_f^{Lev}(\theta)_\perp$ does not exist (as a nonzero matrix). Regarding the Dif moments, when $T>3$, the rank of the orthogonal complement of $A_f^{Dif}(\theta)$, $A_f^{Dif}(\theta)_\perp$, is larger than zero. However, since $\mu_f^{Dif}(\theta,\bar{\sigma}^2)$ equals zero for any value of T, $A_f^{Dif}(\theta)'_\perp \mu_f^{Dif}(\theta,\bar{\sigma}^2)=0$, so the Dif moment conditions do not identify θ. Summarizing, we have:

Dif:	$\mu_f^{Dif}(\theta,\bar{\sigma}^2)$ is vector of all zeros.	No identification when $T\geq 3$.
Lev:	$A_f^{Lev}(\theta)_\perp$ does not exist.	No identification when $T\geq 3$.

(34)

NL condition.

The NL moment condition is not defined for $T=3$. When $T=4$, the expressions of $\mu_f^j(\theta,\bar{\sigma}^2)$, $A_f^j(\theta)$, and $A_f^j(\theta)_\perp$ read

$$\mathbf{NL:} \quad \mu_f^{NL}(\theta,\bar{\sigma}^2)=(1-\theta)\left(\sigma_3^2-\theta\sigma_2^2\right), A_f^{NL}(\theta)=\begin{pmatrix}\theta(\theta-1) & 1-\theta & 0\end{pmatrix},$$
$$A_f^{NL}(\theta)_\perp \text{ does not exist.} \tag{35}$$

Since the orthogonal complement does not exist, the NL moment condition does not identify θ. The expression of $A_f^{NL}(\theta)$ for a larger number of time series observations (see the Appendix) is also such that the orthogonal complement $A_f^{NL}(\theta)_\perp$ also does not exist. Hence, for larger values of T, the NL moment conditions also do not identify θ.

AS and Sys conditions.

The expressions of $\mu_f^j(\theta,\bar{\sigma}^2)$, $A_f^j(\theta)$, and $A_f^j(\theta)_\perp$ when $T=3$ and 4 for the AS and Sys moment conditions result from stacking those of the Dif and NL and Dif and Lev moment conditions, respectively:

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$\textbf{AS:}\quad \mathbf{T=4}\quad \mu_f^{AS}(\theta,\bar{\sigma}^2)=\begin{pmatrix}0\\0\\0\\(1-\theta)\left(\sigma_3^2-\theta\sigma_2^2\right)\end{pmatrix},$$

$$A_f^{AS}(\theta)=\begin{pmatrix}-\theta & 1 & 0\\ 0 & -\theta & 1\\ 0 & -\theta & 1\\ \theta(\theta-1) & 1-\theta & 0\end{pmatrix},$$

$$A_f^{AS}(\theta)_\perp=\begin{pmatrix}\theta-1 & 0\\ 0 & -1\\ 0 & 1\\ 1 & 0\end{pmatrix}.$$

$$\textbf{Sys:}\quad \mathbf{T=3}\quad \mu_f^{Sys}(\theta,\bar{\sigma}^2)=(1-\theta)\begin{pmatrix}0\\ \sigma_2^2\end{pmatrix},A_f^{Sys}(\theta)=\begin{pmatrix}-\theta & 1\\ 1-\theta & 0\end{pmatrix},$$

$A_f^{Sys}(\theta)_\perp$ does not exist.

$$\textbf{Sys:}\quad \mathbf{T=4}\quad \mu_f^{Sys}(\theta,\bar{\sigma}^2)=(1-\theta)\begin{pmatrix}0\\0\\0\\ \sigma_2^2\\ \sigma_3^2\end{pmatrix},A_f^{Sys}(\theta)=\begin{pmatrix}-\theta & 1 & 0\\ 0 & -\theta & 1\\ 0 & -\theta & 1\\ 1-\theta & 0 & 0\\ 0 & 1-\theta & 0\end{pmatrix},$$

$$A_f^{Sys}(\theta)_\perp=\begin{pmatrix}\theta-1 & 0\\ 0 & -1\\ 0 & 1\\ -\theta & 0\\ 1 & 0\end{pmatrix}. \tag{36}$$

When $T=3$, $A_f^{Sys}(\theta)$ is a full-rank square matrix, so its orthogonal complement does not exist. It implies that the Sys moment conditions do not identify θ when $T=3$. When $T=4$, the orthogonal complement of $A_f^j(\theta)$, $A_f^j(\theta)_\perp$, has rank larger than zero for both AS and Sys moments. Furthermore, the specification of $\mu_f^j(\theta,\bar{\sigma}^2)$ for the AS and Sys moment conditions in (36) is such that $A_f^j(\theta)'_\perp\mu_f^j(\theta,\bar{\sigma}^2)\neq 0$, for all $\theta\neq\theta_{0,N}$, while it is not difficult to see that $\lim_{N\to\infty}A_f^j(\theta_{0,N})'_\perp\mu_f^j(\theta_{0,N},\bar{\sigma}^2)=0$ which just reflects that the moment conditions hold at the true value. This implies that although the AS and Sys sample moments diverge in the direction of $A_f^j(\theta)$, so that part cannot be used to identify θ, the AS and Sys sample moments identify θ by their part which is spanned by the orthogonal complement of $A_f^j(\theta)$. The expressions of $\mu_f^j(\theta,\bar{\sigma}^2)$ and $A_f^j(\theta)$ in the proof of Theorem 1 in the Appendix show that this argument extends to all values of T larger than 3.

Our preceding analysis is summarized by Corollary 1.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

COROLLARY 1 (Identification of θ). *Under Assumptions 1 and 2(a), θ is identified by the AS and Sys moment conditions when T exceeds 3. Furthermore, θ is not identified by the Dif, Lev, and NL moment conditions separately for any value of T and the Sys moment conditions when T equals 3.*

Corollary 1 proves Stylized Facts 1–4 from Section 3, which are illustrated by Figures 1 and 2. It also shows that the identification from the Lev moment condition remains problematic for larger values of T, but the Sys and AS moment conditions generally identify θ for values of T larger than 3.

Regarding the NL moments, we find that they are not robust to all settings of nuisance parameters like the variance of the initial observations. Alvarez and Arellano (2004) and Kruiniger (2013) have shown that, when the data, including the initial observation, have finite second moments and the autoregressive parameter equals one, θ is identified by the NL and, hence, the AS moment conditions if and only if $T \geq 4$. Furthermore, if $T \geq 4$, θ is only locally identified when the unconditional variances of the errors change at a constant rate of growth between $t = 2$ and $t = T - 1$ and only second-order but globally identified when the unconditional variances between $t = 2$ and $t = T - 1$ are equal. Unlike Alvarez and Arellano (2004) and Kruiniger (2013), our limiting sequence for the variance of the initial observations allows for unbounded values. Theorem 1 then shows that identification by the NL moment conditions is lost when its convergence rate accords with (23). The intuition is that the NL moment conditions are a product of levels and first differences, so they are unlikely to identify the parameters in limit sequences where the variance of the initial observations increases faster than the sample size.

Theorem 1 can be used to construct the nonstandard limiting behavior of one-step and two-step GMM estimators that result from the different moment conditions. These are similar to the nonstandard results in, e.g., Madsen (2003) and Kruiniger (2009), so we, for reasons of brevity, refrain from stating them.

Robust sample moments

Theorem 1 shows that the identification of θ when the variance of the initial observations is large results from the part of the (AS or Sys) moment conditions that lies in the direction of $A_f^j(\theta)_\perp$. Expressions of the orthogonal complements of $A_f^j(\theta)$ for $T = 4$ and 5 for the AS and Sys moment conditions are stated in (36). They can be specified (see the Appendix) as

$$A_f^j(\theta)_\perp = (G_{f,T}^j(\theta) \vdots G_{2,T}^j), \tag{37}$$

where T indicates the number of time periods and $G_{2,T}^j$ is such that $G_{2,T}^{j\prime} \mu_f^j(\theta, \bar{\sigma}^2) = 0$, for all θ. Furthermore, $G_{f,T}^j(\theta)$ is the only part of $A_f^j(\theta)_\perp$ that depends on θ. The orthogonal complements are then such that the resulting, what we refer to as, robust

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

moment conditions are quadratic in θ:

$$g^j_{f,T}(\theta) = A_f(\theta)^{j'}_{\perp} f^j_N(\theta) = a\theta^2 + b\theta + d, \tag{38}$$

where the expressions for a, b, and d are constructed in the Appendix:

T=4:

Sys: $a = \frac{1}{N}\sum_{i=1}^N \binom{(\Delta y_{i2})^2}{0}$, $b = -\frac{1}{N}\sum_{i=1}^N \binom{(y_{i3}-y_{i1})^2}{\Delta y_{i2}\Delta y_{i3}}$, $d = \frac{1}{N}\sum_{i=1}^N \binom{(y_{i4}-y_{i1})\Delta y_{i3}}{\Delta y_{i2}\Delta y_{i4}}$.

AS: $a = \frac{1}{N}\sum_{i=1}^N \binom{(y_{i3}-y_{i1})\Delta y_{i2}}{0}$, $b = -\frac{1}{N}\sum_{i=1}^N \binom{(y_{i3}-y_{i1})\Delta y_{i3}+(y_{i4}-y_{i1})\Delta y_{i2}}{\Delta y_{i2}\Delta y_{i3}}$, $d = \frac{1}{N}\sum_{i=1}^N \binom{(y_{i4}-y_{i1})\Delta y_{i3}}{\Delta y_{i2}\Delta y_{i4}}$.

T=5:

$$\text{Sys: } a = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (\Delta y_{i2})^2 \\ (y_{i3}-y_{i1})\Delta y_{i3} \\ (\Delta y_{i3})^2 \\ 0 \\ 0 \end{pmatrix},\ b = -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (y_{i3}-y_{i1})^2 \\ (y_{i4}-y_{i1})(y_{i4}-y_{i2}) \\ (y_{i4}-y_{i2})^2 \\ \Delta y_{i2}\Delta y_{i4} \\ \Delta y_{i3}\Delta y_{i4} \end{pmatrix},$$

$$d = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i3} \\ (y_{i5}-y_{i1})\Delta y_{i4} \\ (y_{i5}-y_{i2})\Delta y_{i4} \\ \Delta y_{i2}\Delta y_{i5} \\ \Delta y_{i3}\Delta y_{i5} \end{pmatrix}.$$

$$\text{AS: } a = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (y_{i3}-y_{i1})\Delta y_{i2} \\ (y_{i4}-y_{i1})\Delta y_{i3} \\ (y_{i4}-y_{i2})\Delta y_{i3} \\ 0 \\ 0 \end{pmatrix},$$

$$b = -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i2}+(y_{i3}-y_{i1})\Delta y_{i3} \\ (y_{i4}-y_{i1})\Delta y_{i4}+(y_{i5}-y_{i1})\Delta y_{i3} \\ (y_{i4}-y_{i2})\Delta y_{i4}+(y_{i5}-y_{i2})\Delta y_{i3} \\ \Delta y_{i2}\Delta y_{i4} \\ \Delta y_{i3}\Delta y_{i4} \end{pmatrix},$$

$$d = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i3} \\ (y_{i5}-y_{i1})\Delta y_{i4} \\ (y_{i5}-y_{i2})\Delta y_{i4} \\ \Delta y_{i2}\Delta y_{i5} \\ \Delta y_{i3}\Delta y_{i5} \end{pmatrix},$$

and similar specifications of a, b, and d result for larger values of T.

It is interesting to see that these robust moments only depend on differences of the data, so the initial observations get differenced out. This explains why these moments are robust to the variance of the initial observations. When the autoregressive parameter equals one and in the case of i.i.d. normal errors and time series homoskedasticity, Ahn and Thomas (2006) and Kruiniger (2013)

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

show that the maximum likelihood estimator of Hsiao, Pesaran, and Tahmiscioglu (2002) and the random effects estimator of Anderson and Hsiao (1982) have the same limiting distributions. These results show that, similar to our findings, moment conditions involving levels of the data are redundant in this setting, and only moment conditions using differences of the data, like our robust moment conditions, are informative.

Large individual effect variance.

So far, we have focused on highly persistent panel data resulting from a large autoregressive parameter. However, the representation theorem for the moment conditions and their derivatives in Theorem 1 applies to any setting where the variance of the initial observations gets large. The expression of the initial observation in (7) shows that its variance becomes large when either the variance of the initial disturbance term, u_{i1}, or the individual specific effect, μ_i, becomes large. Theorem 1 focuses on a large variance that results from the autoregressive parameter converging to one. Theorem 1 does, however, extend to the case where jointly with the sample size, the individual specific effect variance becomes large in such a manner that Assumption 2(a) holds. This drifting sequence applies to any value of the autoregressive parameter, so the resulting identification issues are then no longer confined to the unit root value. Hence, they also apply to the cases with only moderate autoregressive dynamics, but a large variance of the unobserved heterogeneity. The robust moments in (38) also apply to this case. Kruiniger (2002) extensively analyzes the setting of a large variance of the individual specific effects. He shows that only moment conditions based on differences of the data yield a consistent estimator, so moment conditions involving levels are redundant. He also constructs the set of optimal moment conditions assuming time series homoskedasticity. Our robust moments (38) extend his set of optimal moment conditions, since they remain valid under a large variance of the individual specific effect and also allow for time series heteroskedasticity.

4. KLM TEST AND ROBUST SAMPLE MOMENTS

Theorem 1 establishes identification results for the AS and Sys moment conditions, which are based on the robust sample moments. It is not clear, however, how an identification robust test procedure makes use of it. In this section, we show that the KLM test based on the original AS or Sys moment conditions just uses the robust sample moments when only the latter contain identifying information on the autoregressive parameter. We show that, under large variances of the initial observation and when the true value of θ is close to one, the KLM test based on either the AS or Sys moment conditions exploits the identifying information from the robust moment conditions in an optimal manner. For practical purposes, this implies that we do not have to explicitly use the robust sample moments, since

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

they are implicitly used when conducting a KLM test using AS or Sys moment conditions.

We obtain the above result in four steps. First, we characterize the limit behavior of the robust sample moments. Second, we use it to determine asymptotic sequences for the true and hypothesized values, so the power properties of the corresponding identification robust test statistics when using the robust moments are not trivial and stay informative. Third, we construct the largest (infeasible) discriminatory power that can be obtained from combining the robust moments. Finally, we show that it coincides with the rejection frequency of KLM tests using either AS or Sys moment conditions. Summarizing, the KLM test based on original AS or Sys moment conditions implicitly resorts to using the robust sample moments in an optimal manner when only these contain information on θ.

4.1. Large Sample Behavior of Robust Sample Moments

To construct the limiting behavior of the robust sample moments for settings where only they contain information on θ, we first state the probability limits of the quantities a, b, and d in (38) under Assumption 1. The components that comprise the robust sample moments do not depend on the variance of the initial observations, so they are not affected by Assumption 2. Since we analyze the behavior when the true value $\theta_{0,N}$ is converging to one, we specify this convergence behavior of $\theta_{0,N}$, so it is dominated by the random components present in the limit behavior of a, b, and d which are of order $O_p(N^{-\frac{1}{2}})$. This then implies that $\theta_{0,N}$ converges rather rapidly to one with a convergence rate that is faster than $N^{-\frac{1}{2}}$. Hence, $\theta_{0,N}$ is considered to be in the close neighborhood of one.

THEOREM 2. *Under Assumption 1, the limit behavior of the different components of $g^j_{f,T}(\theta)$, $j = AS,\ Sys$, for $\theta_{0,N} = 1 + \frac{l}{N^\tau}$ with l a fixed constant, $l < 0$, and $\tau > \frac{1}{2}$, is characterized by:*

$$\mathbf{T} = \mathbf{4} : a = \begin{pmatrix} \sigma_2^2 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),\ b = -\begin{pmatrix} \sigma_2^2 + \sigma_3^2 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}), d = \begin{pmatrix} \sigma_3^2 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}).$$

$$\mathbf{T} = \mathbf{5} : a = \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \\ \sigma_3^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),\ b = \begin{pmatrix} \sigma_2^2 + \sigma_3^2 \\ \sigma_3^2 + \sigma_4^2 \\ \sigma_3^2 + \sigma_4^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),$$

$$d = \begin{pmatrix} \sigma_3^2 \\ \sigma_4^2 \\ \sigma_4^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}).$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Proof. See the Appendix. □

Although AS and Sys robust moments are different, Theorem 2 implies that under Assumption 1, the probability limits of a, b, and d are identical. Furthermore, Theorem 2 implies that the Jacobian of the robust moment equation (38) is of full column rank when $\sigma_t^2 \neq \sigma^2$ for at least one value of $t = 2, \ldots, T$. This fulfills one of the sufficient conditions for standard asymptotic theory for GMM inference based on the robust sample moments, which, since the other sufficient conditions can be shown to hold as well, applies for these settings.

4.2. Asymptotic Sequence for the Hypothesized Value

We want to compare tests of $H_0 : \theta = \theta^*$ using the robust sample moments to KLM tests of H_0 using the original AS and Sys moments for settings where the identification can be problematic, which occurred for true values of θ close to one and large variances of the initial observations. Because we want to analyze local asymptotic power while the true value $\theta_{0,N}$ is converging to one according to $\theta_{0,N} = 1 + \frac{l}{N^\tau}$, we also consider a local to unity drifting sequence for the hypothesized value θ^*, which we denote by $\theta(e)$ with $e < 0$ the localizing parameter. Although less common in asymptotic power analysis, the advantage of a drifting hypothesized value is that our results hold for a range of hypothesized values.

The asymptotic sequence $\theta(e)$ is such that the behavior of the identification robust tests is not diverging and informative about θ, when the true value $\theta_{0,N}$ is converging to one. Theorem 3 establishes the particular rate at which $\theta(e)$ converges to one which makes these conditions hold. Note that there is a slight abuse of notation, as, from now on, we suppress the superscript j in $g^j_{f,T}(\theta(e))$, $j = AS, Sys$, which is inconsequential for the results to follow.

THEOREM 3. *Under Assumption 1, $\theta_{0,N} = 1 + \frac{l}{N^\tau}$ with l a fixed constant, $l < 0$, and $\tau > \frac{1}{2}$, the robust moments $\sqrt{N} g_{f,T}(\theta(e))$ are informative about θ and converge to a bounded in probability, nondegenerate random variable under the following local to unity drifting sequence $\theta(e)$:*

1. $\theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ *in the case of* $\sigma_t^2 = \sigma^2$, $t = 2, \ldots T$,
2. $\theta(e) = 1 + \frac{e}{\sqrt{N}}$ *when* $\sigma_t^2 \neq \sigma^2$, *for at least one value of* t, $t = 2, \ldots T - 1$,

with $e < 0$ a finite constant.

Proof. See the Appendix. □

The quartic root convergence rate in Theorem 3.1 results, since the Jacobian of the robust moment equation (38) is then equal to zero, but the Hessian is not. It is thus a setting of so-called second-order identification with first-order underidentification. Estimators then generally have quartic root convergence rates (see, e.g., Dovonon and Renault, 2013; Dovonon and Hall, 2018; Dovonon et al.,

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

2020). A quartic root convergence rate for estimators in dynamic panel data models is also found by Ahn and Thomas (2006) and Kruiniger (2013).

The quartic root convergence rate for the robust sample moments results from specifying $\theta(e) = 1 + \frac{e}{N^{1/4}}$ and $\sigma_t^2 = \sigma^2,\ t = 2, \dots T$. All elements of the robust sample moments which are linear in e then cancel out in the limit. We are then left with a quadratic term in e and components that converge at the rate $\frac{1}{\sqrt{N}}$. A quartic root convergence rate makes all these components of the same order of magnitude. Theorem 3 shows that error variances which are constant over time, $\sigma_t^2 = \sigma^2,\ t = 2, \dots T,$ lead to this slow convergence rate.

4.3. Largest Rejection Frequencies of Robust Sample Moments

To show that the KLM test of H_0 using AS and Sys moment conditions just uses the robust sample moments when only these contain information on θ, we use the largest rejection frequencies that result in such instances from the robust sample moments. To obtain these largest rejection frequencies, we first consider the GMM-AR test of $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ using the robust sample moments, which is specified as:

$$\text{GMM-AR}(\theta(e)) = N g_{f,T}(\theta(e))' \hat{V}_{gg}(\theta(e))^{-1} g_{f,T}(\theta(e)), \tag{39}$$

with $g_{f,T}(\theta(e))$ the moments in (38) evaluated at $\theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ and $\hat{V}_{gg}(\theta(e))$ the (Eicker–White) covariance matrix estimator of the covariance matrix of $g_{f,T}(\theta(e))$. For $T = 4$ and 5:[7]

$$\mathbf{T = 4}: \quad g^{AS}_{f,T=4}(\theta(e)) = \begin{pmatrix} 1 & -\theta(e) \\ 0 & 1 \end{pmatrix} g^{Sys}_{f,T=4}(\theta(e)),$$

$$\mathbf{T = 5}: \quad g^{AS}_{f,T=5}(\theta(e))$$

$$= \begin{pmatrix} 1 & -\theta(e)/(1-\theta(e)) & \theta(e)/(1-\theta(e)) & 0 & 0 \\ 0 & 1 & 0 & 0 & -\theta(e) \\ 0 & 0 & 1 & 0 & -\theta(e) \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} g^{Sys}_{f,T=5}(\theta(e)),$$

so GMM-AR($\theta(e)$) is equivalent for the AS and Sys moment conditions, since the invertible matrix by which $g^{Sys}_{f,T}(\theta(e))$ has to be premultiplied to obtain $g^{AS}_{f,T}(\theta(e))$ cancels out in GMM-AR($\theta(e)$). This result can be extended to larger values of T.

THEOREM 4. *Under Assumption 1, $\theta_{0,N} = 1 + \frac{l}{N^\tau}$ with l a fixed constant, $l < 0$, and $\tau > \frac{1}{2}$, $\sigma_t^2 = \sigma^2$, $t = 2, \dots, T$, the large sample distribution of the GMM-AR statistic (39) for testing $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$, in a sample of size N is characterized*

[7] We thank an anonymous referee for showing this.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

by

$$\chi^2(\delta(N), p_{\max}), \tag{40}$$

with $\delta(N) = (e\sigma)^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-1} \binom{\iota_p}{0}$, *p the number of columns* $G_{f,T}(\theta)$, *so when* $T = 4$, $p = 1$, *and when* $T = 5$, $p = 3$, *and* $p_{\max}$ *the number of elements of* $g_{f,T}(\theta(e))$, *so, when* $T = 4$, $p_{\max} = 2$, *while* $p_{\max} = 5$, *for* $T = 5$,

$$B(N) = (\iota_3 \otimes I_{p_{\max}}) + \tfrac{e}{\sqrt[4]{N}} \left[(2 + \tfrac{e}{\sqrt[4]{N}})(e_{1,3} \otimes I_{p_{\max}}) + (e_{2,3} \otimes I_{p_{\max}}) \right], \tag{41}$$

V_{abd} *the covariance matrix of a, b, and d,* $I_{p_{\max}}$ *the* $p_{\max} \times p_{\max}$ *dimensional identity matrix,* $e_{1,3}$ *and* $e_{2,3}$ *the first and second* 3×1 *dimensional unity vectors, and* $\chi^2(\delta, p_{\max})$ *a noncentral* χ^2 *distribution with noncentrality parameter* δ *and* $p_{\max}$ *degrees of freedom.*

Proof. See the Appendix. □

The expression of the large sample distribution in Theorem 4 depends on the sample size. Given the quartic root convergence rate, convergence to the limiting distribution is very slow, so it is important for the accuracy of the approximation of the finite sample distribution to incorporate higher-order components. The proof of Theorem 4 in the Appendix, therefore, from the outset considers all higher-order components of $g_{f,T}(\theta(e))$ in order to construct a large sample approximation of the distribution of GMM-AR$(\theta(e))$.

To obtain the maximal rejection frequencies using the robust sample moments, we use a (infeasible) weighted average of the moment equations in $g_{f,T}(\theta(e))$ where the weights are chosen such that the noncentrality parameter equals the one of the noncentral χ^2 limiting distribution of the GMM-AR statistic while the degrees of freedom is equal to one (i.e., the number of elements of θ). This value of the noncentrality parameter is also the maximal one that can be obtained using a weighted average of the robust sample moments.

THEOREM 5. *Under Assumption 1,* $\theta_{0,N} = 1 + \frac{l}{N^\tau}$ *with l a fixed constant,* $l < 0$, *and* $\tau > \frac{1}{2}$, $\sigma_t^2 = \sigma^2$, $t = 2, \ldots T$, *an optimal (infeasible) GMM-AR test of* H_p : $\theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ *that uses a weighted average of the robust sample moments can be constructed that has approximately a*

$$\chi^2(\delta(N), 1), \tag{42}$$

distribution in large samples of size N.

Proof. See the Appendix. □

The GMM-AR statistics in Theorems 4 and 5 both have noncentral χ^2 distributions with the same noncentrality parameter, so the one with the smallest number of degrees of freedom, i.e., the statistic in Theorem 5, has the largest power.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

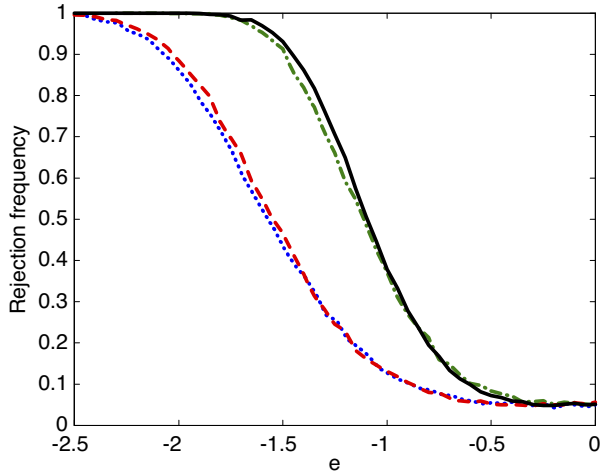


FIGURE 4. Rejection frequencies of GMM-AR tests of $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ using weighted robust sample moments. *Notes:* 5% significance level, true value of θ is 0.99, $N = 2,000$, Sys & $T = 4$ (dashed), AS & $T = 4$ (dotted), Sys & $T = 5$ (solid), and AS & $T = 5$ (dash-dotted).

Figure 4 illustrates Theorem 5 and shows the maximal rejection frequencies based on combining the robust sample moments based on either AS or Sys moment condition in a GMM-AR test[8] for $T = 4$ and 5. It uses DGP 1 from Section 3 with a true value of θ which is very close to one (0.99) and a large value of σ_c^2 (10) compared to σ^2 (one), which amplifies the variance of the initial conditions. The DGP thus satisfies mean stationarity (7)–(9) and also time series homoskedasticity, i.e., $\sigma_t^2 = \sigma^2$, for $t = 2, \ldots, T$. We use $N = 2,000$, a relatively large value and test for a wide range of values for θ, which together with N provides a mapping to the constant e $(= \sqrt[4]{N}(\theta - 1))$ in Figure 4 (horizontal axis). The usual power curve, as shown earlier in Figures 1 and 2, reports the rejection frequencies of tests of the hypothesized parameter value as a function of the parameter value used in the DGP where the data are simulated from. Figure 4, however, reports for a fixed parameter value equal to one in the DGP used to simulate the data, the rejection frequencies as a function of a varying localizing parameter e, and, hence, autoregressive parameter $\theta(e)$, under the tested null hypothesis. The rejection frequencies in Figure 4, thus, report those observed at one for a range of the usual power curves where the tested parameter values correspond with those on the horizontal axes in Figure 4.

Because of the equivalence of the GMM-AR test for the AS and Sys robust moments, the rejection frequencies are identical for the AS- and Sys-based robust

[8]We use the covariance matrix estimator for each simulated dataset to compute the GMM-AR statistics.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

sample moments and only differ over T. Any remaining differences in Figure 4 are due to sampling noise.

4.4. Large Sample Behavior of the KLM Test

Finally, we construct the large sample distribution of KLM tests of $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ using AS and Sys moment conditions when $\theta_{0,N}$ accords with the drifting sequences in Assumptions 1 and 2(a), so only the robust sample moments contain information on θ.

THEOREM 6. *Under Assumptions 1 and 2(a), $\theta_{0,N} = 1 + \frac{l}{N^\tau}$ with l a fixed constant, $l < 0$, and $\tau > \frac{1}{2}$, $\sigma_t^2 = \sigma^2$, $t = 2, \ldots T$, the large sample distribution of the KLM statistic using the AS or Sys moments for testing the hypothesis $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ is characterized by*

$$KLM(\theta(e)) \sim \quad \chi^2(\delta(N), 1), \tag{43}$$

with $\delta(N)$ defined in Theorem 4.

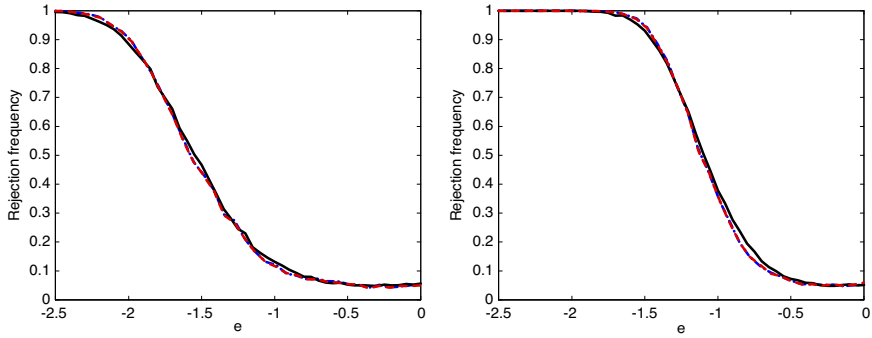
Proof. See the Appendix. □

Under Assumptions 1 and 2(a), Theorem 1 implies that the GMM sample moments diverge in one direction and converge in another one. Identical to tests for cointegration, Theorem 6 shows that the diverging parts of the GMM sample moments cancel out in the large sample distribution of the KLM test, so it only contains elements from the converging part of the GMM sample moments. The proof of the large sample distribution of the KLM test is, therefore, rather elaborate, since this has to be shown for each of the different components of the KLM test.

Theorem 6 shows that the large sample distribution of the KLM test using AS or Sys moment conditions when only the robust sample moments contain information on θ is identical to the limiting distribution of the GMM-AR test that optimally combines the robust sample moments for these settings. It proves that KLM tests using the AS and Sys moment conditions then only use the robust sample moments. It is similar to what happens in cointegration where, since the cointegrating vector and stochastic trends operate orthogonally, a likelihood ratio test on the cointegration vector also does not depend on the stochastic trends (see, e.g., Johansen, 1991).

Theorem 6 is illustrated by Figure 5a,b, which shows the rejection frequencies of 5% significance tests using a KLM test of $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ with AS and Sys moment conditions when T equals 4 (Figure 5a), and 5 (Figure 5b), respectively. It uses the same DGP as for Figure 4. In addition, identical to Figure 4, the rejection frequencies in Figure 5 report the rejection frequencies when using a fixed parameter value in the DGP where we simulate the data from, as a function of a varying parameter value under the tested hypothesis.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press



(a) $T = 4$. (b) $T = 5$.

FIGURE 5. Rejection frequencies of KLM tests of $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ using AS (dashed) and Sys (dash-dotted) and GMM-AR tests using (infeasible) optimal weighted average of robust sample moments (solid line). *Notes:* 5% significance level, true value of θ is 0.99, and $N = 2,000$.

Figure 5 shows, for both $T = 4$ and $T = 5$, that the rejection frequencies that result from using the KLM test with either AS or Sys moment conditions are equal to the largest rejection frequencies, that can be obtained with the robust moments when only they contain information on θ. It illustrates that the robust sample moments are (implicitly) used when you conduct KLM tests with AS or Sys moment conditions. Hence, in practice, one can just use AS or Sys moment conditions in the construction of the KLM test, i.e., there is no need to switch to the robust sample moments.

Figure 5 also provides a visual proof of Stylized Fact 5 from Section 3, i.e., rejection frequencies for the KLM test using AS or Sys moment conditions are almost identical when the true value of θ is close to one and for large variances of the initial observations, and that it is not specific for the tested values used there but holds generally for different tested values of θ.

5. CONCLUSIONS

We have analyzed GMM inference for dynamic panel data models involving highly persistent panel data. We show that the Dif, Lev, and NL moment conditions separately do not identify the parameters in dynamic panel data models for a general number of time periods. This results from the divergence of the initial observations for some plausible DGP involving highly persistent panel data. When there are more than three time periods, the AS and Sys moment conditions, however, do lead to identification. The identification based on the AS and Sys moment conditions for the problematic cases of divergent initial observations results from so-called robust sample moments. They are combinations of either the AS or Sys sample moments and do not depend on the initial observations.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Despite the positive identification results for AS and Sys moment conditions, conventional inference based on two-step GMM estimators is not valid, since these estimators have nonstandard limiting distributions near the unit root. Similar results hold for two-step GMM estimators based on our robust sample moments. We have, therefore, analyzed the large sample properties of identification robust GMM test procedures. These test statistics are size correct, easy to implement, and have been used in a variety of models analyzed using GMM. We show that the identification robust KLM statistic based on the AS and Sys sample moments implicitly resorts to using the robust sample moments when only the latter contain identifying information.

Based on the theoretical analysis and numerical results, a number of remarks can be made regarding the implementation of GMM inference for applied linear dynamic panel data analysis. First, statistical inference, i.e., hypothesis testing and confidence intervals, should be based on identification robust tests, like the KLM or GMM-AR test. The nonstandard limiting behavior of the two-step GMM coefficient estimator makes the use of conventional GMM inference hazardous in applied research when there are identification issues. Second, one should always use either AS or Sys moment conditions, since these deliver identification under more general conditions when $T > 3$. An advantage of the AS moments is that they are valid under less restrictive assumptions than the Sys moments. Third, when mean stationarity applies, the Sys moments are preferred. Although AS and Sys moments contain the same amount of identifying information when θ is close to one and the variance of the initial observations is large, in practice, the opposite may well be the case if one is not close to the unit root (or if time series heteroskedasticity is present). This is shown, for example, by our simulated KLM power curves in Section 2. Fourth, the original AS or Sys moments should be used in an identification robust GMM test statistic and not the implied robust sample moments. Although only the latter preserve identification when the variance of the initial observations is large, we have shown that the identification robust KLM test based on the AS or Sys moments implicitly uses the robust sample moments.

Finally, for expository purposes, we have only analyzed the first-order autoregressive panel data model. The extension to panel data models with multiple endogenous regressors, e.g., dynamic models with additional endogenous regressors, is an important area for future research.

APPENDICES

A. Specification of GMM Sample Moments and Proofs

Specification of sample moment functions. For the Dif moment conditions in (4), k_{Dif} equals $\frac{1}{2}(T-2)(T-1)$ while $f_i^{Dif}(\theta)$ and $q_i^{Dif}(\theta)$ read

$$f_i^{Dif}(\theta) = Z_i^{Dif} \varphi_i^{Dif}(\theta),$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$q_i^{Dif}(\theta) = -Z_i^{Dif}\,\Delta y_{-1,i},$$

with $\varphi_i^{Dif}(\theta) = (\Delta y_{i3} - \theta\Delta y_{i2} \dots \Delta y_{iT} - \theta\Delta y_{iT-1})'$, $\Delta y_{-1,i} = (\Delta y_{i2} \dots \Delta y_{iT-1})'$ and

$$Z_i^{Dif} = \begin{pmatrix} y_{i1} & 0\dots 0 & 0 \\ 0 & \ddots & 0 \\ & & \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT-2} \end{pmatrix} \\ 0 & 0\dots 0 & \end{pmatrix} : \frac{1}{2}(T-1)(T-2) \times (T-2).$$

For the Lev moment conditions in (5), k_{Lev} equals $T-2$ while the sample moment functions are

$$f_i^{Lev}(\theta) = Z_i^{Lev}\varphi_i^{Lev}(\theta),$$
$$q_i^{Lev}(\theta) = -Z_i^{Lev}y_{-1,i},$$

with $\varphi_i^{Lev}(\theta) = (y_{i3} - \theta y_{i2} \dots y_{iT} - \theta y_{iT-1})'$, $y_{-1,i} = (y_{i2} \dots y_{iT-1})'$, and

$$Z_i^{Lev} = \begin{pmatrix} \Delta y_{i2} & 0\dots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\dots 0 & \Delta y_{iT-1} \end{pmatrix} : (T-2) \times (T-2).$$

For the NL moment conditions in (10), k_{NL} equals $T-3$ while the sample moment functions can be specified as

$$f_i^{NL}(\theta) = Z_i^{NL}(\theta)\varphi_i^{NL}(\theta),$$
$$q_i^{NL}(\theta) = \left(\frac{\partial}{\partial\theta}Z_i^{NL}(\theta)\right)\varphi_i^{NL}(\theta) + Z_i^{NL}(\theta)\left(\frac{\partial}{\partial\theta}\varphi_i^{NL}(\theta)\right),$$

with $\varphi_i^{NL}(\theta) = ((y_{i4} - \theta y_{i3}) \dots (y_{iT} - \theta y_{iT-1}))'$ and

$$Z_i^{NL}(\theta) = \begin{pmatrix} (\Delta y_{i3} - \theta\Delta y_{i2}) & 0\dots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\dots 0 & (\Delta y_{iT-1} - \theta\Delta y_{iT-2}) \end{pmatrix} : (T-3) \times (T-3).$$

The sample moments for the AS moment conditions result by just stacking the appropriate sample moments stated above, so k_{AS} equals $\frac{1}{2}(T-1)(T-2) + T - 3$. In a similar manner, the Sys sample moments result, so k_{Sys} equals $\frac{1}{2}(T+1)(T-2)$.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

LEMMA 1. *We state some intermediate results, which involve the different terms in the sample moments and their derivatives. Assumption 1 implies the following:*

i.	$\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)y_{i1}u_{i1} =$	$-d_2-\frac{\sigma_{1,N}}{\sqrt{N}}\psi_c+o_p(1),$	
ii.	$\frac{1}{N}\sum_{i=1}^{N}(1-\theta_{0,N})u_{i1}u_{it} \underset{p}{\to}$	$0,$	$t>1,$
iii.	$\frac{1}{N}\sum_{i=1}^{N}u_{it}^2 \underset{p}{\to}$	$\sigma_t^2,$	$t>1,$
iv.	$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{it}\Delta y_{it} \underset{p}{\to}$	$\sigma_t^2,$	$t>1,$
v.	$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{it}\Delta y_{is} \underset{p}{\to}$	$0,$	$t,s>1, t\neq s.$
vi.	$\frac{h_N(\theta_{0,N})}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}u_{i2} \\ \vdots \\ y_{i1}u_{iT}\end{pmatrix} \underset{d}{\to}$	$\psi,$	

with $\psi=(\psi_2\ldots\psi_T)'\sim N(0,\text{diag }(\sigma_2^2,\ldots,\sigma_T^2))$ *independent from* $\psi_c\sim N(0,\sigma_c^2)$, $\sigma_c^2=var(c_i)$.

Proof of Lemma 1. i. Under mean stationarity, we have:

$$\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)y_{i1}u_{i1}=\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}^2+\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}\mu_i.$$

Assumption 1(c) implies that $(1-\theta_{0,N})^{\frac{1}{2}}u_{i1}$ is a random variable with finite fourth moments, so a law of large numbers applies:

$$\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}^2\underset{p}{\to}-d_2.$$

Since $c_i=(1-\theta_{0,N})\mu_i$, we can specify:

$$\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}\mu_i=-\frac{1}{N}\sum_{i=1}^{N}u_{i1}c_i=-\frac{\sigma_{1,N}}{\sqrt{N}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{u_{i1}}{\sigma_{1,N}}c_i,$$

because

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{u_{i1}}{\sigma_{1,N}}c_i\underset{d}{\to}\psi_c,$$

with ψ_c independent of $\psi_j, j=2,\ldots,T$, as c_i is independent from $u_{ij}, j=2,\ldots,T$. Upon combining, we obtain:

$$\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)y_{i1}\mu_i=-d_2-\frac{\sigma_{1,N}}{\sqrt{N}}\psi_c+o_p(1).$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

ii. Since u_{it} are independently distributed, $t = 1, \ldots, T$, and $(1-\theta_{0,N})^{\frac{1}{2}} u_{i1}$ is a random variable with finite fourth moments, a law of large numbers applies:

$$\frac{1}{N}\sum_{i=1}^{N}(1-\theta_{0,N})u_{i1}u_{it} \underset{p}{\to} \quad 0, \quad t > 1.$$

iii. Finite fourth moments of u_{it} imply that a law of large numbers applies:

$$\frac{1}{N}\sum_{i=1}^{N}u_{it}^2 \underset{p}{\to} \quad \sigma_t^2, \quad t > 1.$$

iv. Mean stationarity implies $\Delta y_{i2} = u_{i2} + \left(\theta_{0,N} - 1\right) u_{i1}$, so

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i2} = \frac{1}{N}\sum_{i=1}^{N}u_{i2}^2 + \left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}\left(\theta_{0,N}-1\right)u_{i1}^2 + \frac{2}{N}\sum_{i=1}^{N}\left(\theta_{0,N}-1\right)u_{i1}u_{i2}.$$

Because $\frac{1}{N}\sum_{i=1}^{N}(1-\theta_{0,N})u_{i1}^2 \underset{p}{\to} d_2$ and $(1-\theta_{0,N}) \underset{N\to\infty}{\to} 0$, we have

$$\left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}\left(\theta_{0,N}-1\right)u_{i1}^2 \underset{p}{\to} 0,$$

which shows that $\left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}\left(\theta_{0,N}-1\right)u_{i1}^2 = o_p(1)$. Furthermore, since both $\left(\theta_{0,N}-1\right)^{\frac{1}{2}} u_{i1}$ and u_{i2} have finite fourth moments and are independent, $\frac{2}{N}\sum_{i=1}^{N}\left(\theta_{0,N}-1\right) u_{i1}u_{i2} = o_p(1)$, which implies that

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i2} = \frac{1}{N}\sum_{i=1}^{N}u_{i2}^2 + o_p(1).$$

Finally, we have $E\left(u_{i2}^2\right) = \sigma_2^2$ and finite fourth moments; hence,

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}^2 \underset{p}{\to} \sigma_2^2.$$

Along the same lines as the above, this can be shown to hold for other values of t as well.

v. Similar to the above, when substituting for Δy_{i2} and Δy_{i3}, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i3} = {} & \frac{1}{N}\sum_{i=1}^{N}u_{i2}u_{i3} + \left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}u_{i2}^2 + \theta_{0,N}\left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}u_{i1}u_{i2} \\ & + \left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}u_{i1}u_{i3} + \left(\theta_{0,N}-1\right)^2\frac{1}{N}\sum_{i=1}^{N}u_{i1}u_{i2} \\ & + \theta_{0,N}\left(\theta_{0,N}-1\right)^2\frac{1}{N}\sum_{i=1}^{N}u_{i1}^2.\end{aligned}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Similar derivations as before show that $\frac{1}{N}\sum_{i=1}^{N}\theta_{0,N}\left(\theta_{0,N}-1\right)^2 u_{i1}^2 \underset{p}{\to} 0$, $\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)^2 u_{i1}u_{i2} \underset{p}{\to} 0$, $\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i1}u_{i3} \underset{p}{\to} 0$, $\frac{1}{N}\sum_{i=1}^{N}\theta_{0,N}\left(\theta_{0,N}-1\right)u_{i1}u_{i2} \underset{p}{\to} 0$, $\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)u_{i2}^2 \underset{p}{\to} 0$, $\frac{1}{N}\sum_{i=1}^{N}u_{i2}u_{i3} \underset{p}{\to} 0$, so all these terms are $o_p(1)$ and have probability limit 0, implying that

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i3} \underset{p}{\to} 0.$$

Along similar lines, this can be proved to extend to the first differences at other time periods.

vi. Since $h_N(\theta_{0,N})^{-2} = var(y_{i1})$, the random variable $h_N(\theta_{0,N})y_{i1}$ has variance equal to one. Since y_{i1} and u_{it}, $t > 1$, are independent, because of Assumption 1(e), $E(h_N(\theta_{0,N})y_{i1}u_{it}) = 0$. Furthermore, Assumption 1(d) implies that $Var\left(h_N(\theta_{0,N})y_{i1}u_{it}\right) = \sigma_t^2$, which is finite. A central limit theorem therefore applies:

$$\frac{h_N(\theta_{0,N})}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}u_{i2} \\ \vdots \\ y_{i1}u_{iT}\end{pmatrix} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} h_N(\theta_{0,N})y_{i1}u_{i2} \\ \vdots \\ h_N(\theta_{0,N})y_{i1}u_{iT}\end{pmatrix} \underset{d}{\to} \psi,$$

with $\psi = (\psi_{y_{1i}u_{i2}}\ldots\psi_{y_{1i}u_{iT}})'$ a $(T-1)$-dimensional, mean-zero normal random vector. Assumption 1(e) states that $u_{i1}/\sigma_{1,N}$, $u_{i2}\ldots,u_{iT}$, and c_i are independently distributed within individuals and over the different individuals. It implies that $u_{i1}c_i$ and $y_{i1}u_{it}$ are uncorrelated. Since ψ and ψ_c are the limits of the scaled sums of $y_{i1}u_{it}$ and $u_{i1}c_i$, they are uncorrelated normal random variables and therefore independent. As a result of this, the $T \times T$ covariance matrix of ψ and ψ_c is diagonal:

$$V_{\binom{\psi}{\psi_c}\binom{\psi}{\psi_c}} = var(\psi,\psi_c) = \begin{pmatrix} V_{\psi\psi} & V_{\psi\psi_c} \\ V_{\psi_c\psi} & V_{\psi_c\psi_c}\end{pmatrix}$$

$$= E\left[\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} h_N(\theta_{0,N})y_{i1}u_{i2} \\ \vdots \\ h_N(\theta_{0,N})y_{i1}u_{iT} \\ \frac{u_{i1}}{\sigma_{1,N}}c_i\end{pmatrix}\begin{pmatrix} h_N(\theta_{0,N})y_{i1}u_{i2} \\ \vdots \\ h_N(\theta_{0,N})y_{i1}u_{iT} \\ \frac{u_{i1}}{\sigma_{1,N}}c_i\end{pmatrix}'\right]$$

$$= E\left[\lim_{N\to\infty}\begin{pmatrix} h_N(\theta_{0,N})y_{i1}u_{i2} \\ \vdots \\ h_N(\theta_{0,N})y_{i1}u_{iT} \\ \frac{u_{i1}}{\sigma_{1,N}}c_i\end{pmatrix}\begin{pmatrix} h_N(\theta_{0,N})y_{i1}u_{i2} \\ \vdots \\ h_N(\theta_{0,N})y_{i1}u_{iT} \\ \frac{u_{i1}}{\sigma_{1,N}}c_i\end{pmatrix}'\right]$$

$$= \mathrm{diag}(\sigma_2^2\ldots\sigma_T^2\sigma_c^2).$$

□

Proof of Theorem 1. T = 3. Under mean stationarity, we have

$$\Delta y_{i2} = u_{i2} + \left(\theta_{0,N}-1\right)u_{i1},$$
$$\Delta y_{i3} = u_{i3} + \left(\theta_{0,N}-1\right)u_{i2} + \theta_{0,N}\left(\theta_{0,N}-1\right)u_{i1}.$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Substituting these expressions, we can specify the Dif sample moment and its derivative as

$$\begin{aligned} f_N^{Dif}(\theta) &= \tfrac{1}{N}\sum_{i=1}^{N}(y_{i1}\Delta y_{i3} - \theta y_{i1}\Delta y_{i2}) \\ &= \tfrac{1}{N}\sum_{i=1}^{N} y_{i1}u_{i3} + \left(\theta_{0,N} - 1 - \theta\right)\tfrac{1}{N}\sum_{i=1}^{N} y_{i1}u_{i2} + (\theta_{0,N} - \theta)\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N} - 1)y_{i1}u_{i1}, \\ q_N^{Dif}(\theta) &= -\tfrac{1}{N}\sum_{i=1}^{N} y_{i1}\Delta y_{i2} \\ &= -\tfrac{1}{N}\sum_{i=1}^{N} y_{i1}u_{i2} - \tfrac{1}{N}\sum_{i=1}^{N}(\theta_{0,N} - 1)y_{i1}u_{i1}. \end{aligned}$$

Combining convergence results stated in Lemma 1, the large sample behavior of the Dif sample moment and derivative can thus be characterized by

$$\begin{aligned} f_N^{Dif}(\theta) &= \frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left[(\psi_3 - \theta\psi_2) - (1-\theta)h_N(\theta_{0,N})\sigma_{1,N}\psi_c\right] - (1-\theta)\,d_2 + o_p(1), \\ q_N^{Dif}(\theta) &= -\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left[\psi_2 - h_N(\theta_{0,N})\sigma_{1,N}\psi_c\right] + d_2 + o_p(1), \end{aligned}$$

where we note that $h_N(\theta_{0,N})\sigma_{1,N} \leq 1$, since $\mathrm{var}(y_{i1}) \geq \mathrm{var}(u_{i1})$, from which it is readily seen that

$$\begin{aligned} A_f^{Dif}(\theta) &= \begin{pmatrix} -\theta & 1 \end{pmatrix}, \mu_f^{Dif}(\theta, \bar{\sigma}^2) = 0, \\ A_q^{Dif}(\theta) &= \begin{pmatrix} -1 & 0 \end{pmatrix}, \mu_q^{Dif}(\theta, \bar{\sigma}^2) = 0. \end{aligned}$$

Regarding the Lev moment, using

$$\begin{aligned} y_{i2} &= \Delta y_{i2} + y_{i1}, \\ y_{i3} &= \Delta y_{i3} + \Delta y_{i2} + y_{i1}, \end{aligned}$$

we have

$$\begin{aligned} f_N^{Lev}(\theta) &= \tfrac{1}{N}\sum_{i=1}^{N}(y_{i3} - \theta y_{i2})\Delta y_{i2} \\ &= \tfrac{1}{N}\sum_{i=1}^{N}(\Delta y_{i3} + (1-\theta)\Delta y_{i2})\,\Delta y_{i2} + (1-\theta)\tfrac{1}{N}\sum_{i=1}^{N} y_{i1}\Delta y_{i2}. \end{aligned}$$

Exploiting mean stationarity and substituting for Δy_{i2} and Δy_{i3}, we write

$$(1-\theta)\tfrac{1}{N}\sum_{i=1}^{N} y_{i1}\Delta y_{i2} = (1-\theta)\frac{1}{N}\sum_{i=1}^{N} y_{i1}u_{i2} + (1-\theta)\left(\theta_{0,N} - 1\right)\tfrac{1}{N}\sum_{i=1}^{N} y_{i1}u_{i1},$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

and using Lemma 1, we have

$$\frac{1}{N}\sum_{i=1}^{N}(\Delta y_{i3}+(1-\theta)\Delta y_{i2})\,\Delta y_{i2}=(1-\theta)\,\frac{1}{N}\sum_{i=1}^{N}u_{i2}^2+o_p(1).$$

Regarding the Lev derivative, we have

$$\begin{aligned}q_N^{Lev}(\theta)&=-\frac{1}{N}\sum_{i=1}^{N}y_{i2}\Delta y_{i2}\\&=-\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i2}-\frac{1}{N}\sum_{i=1}^{N}y_{i1}\Delta y_{i2},\end{aligned}$$

where

$$\frac{1}{N}\sum_{i=1}^{N}y_{i1}\Delta y_{i2}=\frac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i2}+\left(\theta_{0,N}-1\right)\frac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i1},$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i2}=\frac{1}{N}\sum_{i=1}^{N}u_{i2}^2+o_p(1).$$

Therefore, we can write the Lev moment condition and derivative as

$$f_N^{Lev}(\theta)=(1-\theta)\left\{\frac{1}{N}\sum_{i=1}^{N}u_{i2}^2+\frac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i2}+\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)y_{i1}u_{i1}\right\}+o_p(1).$$

$$q_N^{Lev}(\theta)=-\frac{1}{N}\sum_{i=1}^{N}u_{i2}^2-\frac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i2}-\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)y_{i1}u_{i1}+o_p(1).$$

Combining this and other convergence results from Lemma 1, the large sample behavior of the Lev sample moment and derivative can thus be characterized by

$$f_N^{Lev}(\theta)=(1-\theta)\left\{\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left[\psi_2-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\right]+\left(\sigma_2^2-d_2\right)\right\}+o_p(1),$$

$$q_N^{Lev}(\theta)=-\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left[\psi_2-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\right]-\left(\sigma_2^2-d_2\right)+o_p(1),$$

so this implies that

$$\begin{aligned}&A_f^{Lev}(\theta)=(1-\theta\ 0),\mu_f^{Lev}(\theta,\bar{\sigma}^2)=(1-\theta)\,\sigma_2^2,\\&A_q^{Lev}(\theta)=(-1\ 0),\mu_q^{Lev}(\theta,\bar{\sigma}^2)=-\sigma_2^2.\end{aligned}$$

From this last result, it is not difficult to see that, under Assumption 2(b), we have

$$\frac{1}{N}\sum_{i=1}^{N}y_{i2}\Delta y_{i2}\underset{p}{\to}\sigma_2^2-d_2.$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

The reason for this is that Assumption 2(b) amounts to $h_N(\theta_{0,N})\sqrt{N} = \frac{\sqrt{N}}{\sqrt{var(y_{i1})}} \underset{N\to\infty}{\to} \infty$, and, since $\text{var}(y_{i1}) \geq \text{var}(u_{i1})$, it implies that $\sigma^2_{1,N}/N \underset{N\to\infty}{\to} 0$. Finally, the Sys sample moment and derivative simply result from stacking the Dif and Lev sample moments and derivatives:

$$f_N^{Sys}(\theta) = \tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}\Delta y_{i3} - \theta y_{i1}\Delta y_{i2} \\ y_{i3}\Delta y_{i2} - \theta y_{i2}\Delta y_{i2} \end{pmatrix},$$
$$q_N^{Sys}(\theta) = -\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}\Delta y_{i2} \\ y_{i2}\Delta y_{i2} \end{pmatrix}.$$

Combining earlier convergence results, the large sample behavior of the Sys sample moment and derivative can thus be characterized by

$$f_N^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 \\ 1-\theta & 0 \end{pmatrix}\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix} - h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_2\right\} - \iota_2 d_2\right]$$
$$+ (1-\theta)\begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix} + o_p(1),$$
$$q_N^{Sys}(\theta) = -\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix} - h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_2\right\} - \iota_2 d_2\right]$$
$$- \begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix} + o_p(1),$$

from which it is readily seen that

$$A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 \\ 1-\theta & 0 \end{pmatrix}, \mu_f^{Sys}(\theta, \bar{\sigma}^2) = (1-\theta)\begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix},$$
$$A_q^{Sys}(\theta) = \begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}, \mu_q^{Sys}(\theta, \bar{\sigma}^2) = \begin{pmatrix} 0 \\ -\sigma_2^2 \end{pmatrix}.$$

T = 4. Under mean stationarity, we have

$$\Delta y_{i2} = u_{i2} + (\theta_{0,N} - 1)u_{i1},$$
$$\Delta y_{i3} = u_{i3} + (\theta_{0,N} - 1)u_{i2} + \theta_{0,N}(\theta_{0,N} - 1)u_{i1},$$
$$\Delta y_{i4} = u_{i4} + (\theta_{0,N} - 1)u_{i3} + \theta_{0,N}(\theta_{0,N} - 1)u_{i2} + \theta^2_{0,N}(\theta_{0,N} - 1)u_{i1}.$$

Substituting these expressions and $y_{i2} = \Delta y_{i2} + y_{i1}$, we can specify the Dif sample moments and their derivatives as

$$f_N^{Dif}(\theta) = \tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}\Delta y_{i3} - \theta y_{i1}\Delta y_{i2} \\ y_{i1}\Delta y_{i4} - \theta y_{i1}\Delta y_{i3} \\ y_{i2}\Delta y_{i4} - \theta y_{i2}\Delta y_{i3} \end{pmatrix}$$
$$= \begin{pmatrix} \theta_{0,N} - 1 - \theta & 1 & 0 \\ (\theta_{0,N} - \theta)(\theta_{0,N} - 1) & \theta_{0,N} - 1 - \theta & 1 \\ (\theta_{0,N} - \theta)(\theta_{0,N} - 1) & \theta_{0,N} - 1 - \theta & 1 \end{pmatrix}\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$+(\theta_{0,N}-\theta)(\theta_{0,N}-1)\begin{pmatrix}1\\ \theta_{0,N}\\ \theta_{0,N}\end{pmatrix}\frac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i1}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}0\\ 0\\ \Delta y_{i2}(\Delta y_{i4}-\theta\Delta y_{i3})\end{pmatrix},$$

$$q_N^{Dif}(\theta)=-\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}y_{i1}\Delta y_{i2}\\ y_{i1}\Delta y_{i3}\\ y_{i2}\Delta y_{i3}\end{pmatrix}$$
$$=-\begin{pmatrix}1&0&0\\ \theta_{0,N}-1&1&0\\ \theta_{0,N}-1&1&0\end{pmatrix}\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\end{pmatrix}-(\theta_{0,N}-1)\begin{pmatrix}1\\ \theta_{0,N}\\ \theta_{0,N}\end{pmatrix}\frac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i1}$$
$$-\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}0\\ 0\\ \Delta y_{i2}\Delta y_{i3}\end{pmatrix}.$$

The limit behavior of the first two terms in each expression has been established before. Furthermore, Lemma 1 shows that the last term in each expression is $o_p(1)$. Therefore, the large Dif sample moment and derivative can be expressed as:

$$f_N^{Dif}(\theta)=\begin{pmatrix}-\theta&1&0\\ 0&-\theta&1\\ 0&-\theta&1\end{pmatrix}\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix}\psi_2\\ \psi_3\\ \psi_4\end{pmatrix}-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_3\right\}-\iota_3 d_2\right]$$
$$+o_p(1),$$
$$q_N^{Dif}(\theta)=-\begin{pmatrix}1&0&0\\ 0&1&0\\ 0&1&0\end{pmatrix}\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix}\psi_2\\ \psi_3\\ \psi_4\end{pmatrix}-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_3\right\}-\iota_3 d_2\right]$$
$$+o_p(1),$$

from which it is readily seen that

$$A_f^{Dif}(\theta)=\begin{pmatrix}-\theta&1&0\\ 0&-\theta&1\\ 0&-\theta&1\end{pmatrix},\mu_f^{Dif}(\theta,\bar{\sigma}^2)=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix},$$

$$A_q^{Dif}(\theta)=-\begin{pmatrix}1&0&0\\ 0&1&0\\ 0&1&0\end{pmatrix},\mu_q^{Dif}(\theta,\bar{\sigma}^2)=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}.$$

After some algebra, we can specify the Lev sample moments and their derivatives as

$$f_N^{Lev}(\theta)=\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}y_{i3}\Delta y_{i2}-\theta y_{i2}\Delta y_{i2}\\ y_{i4}\Delta y_{i3}-\theta y_{i3}\Delta y_{i3}\end{pmatrix}$$
$$=\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}1-\theta&0&0\\ (1-\theta)(\theta_{0,N}-1)&1-\theta&0\end{pmatrix}\begin{pmatrix}y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\end{pmatrix}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$+(1-\theta)(\theta_{0,N}-1)\begin{pmatrix}1\\ \theta_{0,N}\end{pmatrix}\tfrac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i1}$$

$$+(1-\theta)\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\Delta y_{i2}\Delta y_{i2}\\ \Delta y_{i3}\Delta y_{i3}\end{pmatrix}+\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\Delta y_{i3}\Delta y_{i2}\\ (\Delta y_{i4}+(1-\theta)\Delta y_{i2})\,\Delta y_{i3}\end{pmatrix},$$

$$q_N^{Lev}(\theta)=-\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}y_{i2}\Delta y_{i2}\\ y_{i3}\Delta y_{i3}\end{pmatrix}$$

$$=-\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}1 & 0 & 0\\ \theta_{0,N}-1 & 1 & 0\end{pmatrix}\begin{pmatrix}y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\end{pmatrix}-(\theta_{0,N}-1)\begin{pmatrix}1\\ \theta_{0,N}\end{pmatrix}\tfrac{1}{N}\sum_{i=1}^{N}y_{i1}u_{i1}$$

$$-\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\Delta y_{i2}\Delta y_{i2}\\ \Delta y_{i3}\Delta y_{i3}\end{pmatrix}-\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}0\\ \Delta y_{i2}\Delta y_{i3}\end{pmatrix}.$$

Using Lemma 1, the large sample behavior of these expressions is equal to:

$$f_N^{Lev}(\theta)=\begin{pmatrix}1-\theta & 0 & 0\\ 0 & 1-\theta & 0\end{pmatrix}$$

$$\times\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix}\psi_2\\ \psi_3\\ \psi_4\end{pmatrix}-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_2\right\}-\iota_2 d_2\right]$$

$$+(1-\theta)\begin{pmatrix}\sigma_2^2\\ \sigma_3^2\end{pmatrix}+o_p(1),$$

$$q_N^{Lev}(\theta)=-\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix}\psi_2\\ \psi_3\\ \psi_4\end{pmatrix}-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_2\right\}-\iota_2 d_2\right]$$

$$-\begin{pmatrix}\sigma_2^2\\ \sigma_3^2\end{pmatrix}+o_p(1),$$

so this implies that

$$A_f^{Lev}(\theta)=\begin{pmatrix}1-\theta & 0 & 0\\ 0 & 1-\theta & 0\end{pmatrix},\mu_f^{Lev}(\theta,\bar{\sigma}^2)=(1-\theta)\begin{pmatrix}\sigma_2^2\\ \sigma_3^2\end{pmatrix},$$

$$A_q^{Lev}(\theta)=-\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix},\mu_q^{Lev}(\theta,\bar{\sigma}^2)=-\begin{pmatrix}\sigma_2^2\\ \sigma_3^2\end{pmatrix}.$$

We can specify the NL sample moment and its derivative as

$$f_N^{NL}(\theta)=\tfrac{1}{N}\sum_{i=1}^{N}(y_{i4}-\theta y_{i3})\,(\Delta y_{i3}-\theta\Delta y_{i2})$$

$$=\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(1-\theta)(\theta_{0,N}-\theta-1) & (1-\theta) & 0\end{pmatrix}\begin{pmatrix}y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\end{pmatrix}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$
\begin{aligned}
&+\tfrac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)(\theta_{0,N}-\theta)(1-\theta)y_{i1}u_{i1}\\
&+(1-\theta)\,\tfrac{1}{N}\sum_{i=1}^{N}(\Delta y_{i3}\Delta y_{i3}-\theta\Delta y_{i2}\Delta y_{i2})\\
&+\tfrac{1}{N}\sum_{i=1}^{N}((\Delta y_{i4}+(1-\theta)\,\Delta y_{i2})\Delta y_{i3}-(\Delta y_{i4}+(1-\theta)\,\Delta y_{i3})\theta\Delta y_{i2}),\\
q_N^{NL}(\theta) &= -\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\theta_{0,N}-2\theta & -1 & 0\end{pmatrix}\begin{pmatrix}y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\end{pmatrix}\\
&+\tfrac{1}{N}\sum_{i=1}^{N}(\theta_{0,N}-1)(1+\theta_{0,N}-2\theta)y_{i1}u_{i1}\\
&-\tfrac{1}{N}\sum_{i=1}^{N}(\Delta y_{i3}\Delta y_{i3}+(1-2\theta)\,\Delta y_{i2}\Delta y_{i2})\\
&-\tfrac{1}{N}\sum_{i=1}^{N}(\Delta y_{i2}\Delta y_{i3}+(\Delta y_{i4}+(1-2\theta)\,\Delta y_{i3})\,\Delta y_{i2}).
\end{aligned}
$$

Using Lemma 1, the large sample behavior of these expressions is equal to:

$$
\begin{aligned}
f_N^{NL}(\theta) &= \tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\theta(\theta-1) & 1-\theta & 0\end{pmatrix}\\
&\times\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix}\psi_2\\ \psi_3\\ \psi_4\end{pmatrix}-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_2\right\}-\iota_2 d_2\right]\\
&+(1-\theta)\left(\sigma_3^2-\theta\sigma_2^2\right)+o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
q_N^{NL}(\theta) &= -\tfrac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}1-2\theta & 1 & 0\end{pmatrix}\\
&\times\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left\{\begin{pmatrix}\psi_2\\ \psi_3\\ \psi_4\end{pmatrix}-h_N(\theta_{0,N})\sigma_{1,N}\psi_c\iota_2\right\}-\iota_2 d_2\right]\\
&-\sigma_3^2-(1-2\theta)\sigma_2^2+o_p(1),
\end{aligned}
$$

so this implies that:

$$
\begin{aligned}
&A_f^{NL}(\theta)=\begin{pmatrix}\theta(\theta-1) & 1-\theta & 0\end{pmatrix},\quad \mu_f^{NL}(\theta,\bar{\sigma}^2)=(1-\theta)\left(\sigma_3^2-\theta\sigma_2^2\right),\\
&A_q^{NL}(\theta)=\begin{pmatrix}2\theta-1 & -1 & 0\end{pmatrix},\quad \mu_q^{NL}(\theta,\bar{\sigma}^2)=(2\theta-1)\sigma_2^2-\sigma_3^2.
\end{aligned}
$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Finally, regarding AS and Sys moment conditions, we simply have

$$A_f^{Sys}(\theta) = \begin{pmatrix} A_f^{Dif}(\theta) \\ A_f^{Lev}(\theta) \vdots 0 \end{pmatrix}, \quad \mu_f^{Sys}(\theta, \bar{\sigma}^2) = \begin{pmatrix} \mu_f^{Dif}(\theta, \bar{\sigma}^2) \\ \mu_f^{Lev}(\theta, \bar{\sigma}^2) \end{pmatrix},$$

$$A_q^{Sys}(\theta) = \begin{pmatrix} A_q^{Dif}(\theta) \\ A_q^{Lev}(\theta) \vdots 0 \end{pmatrix}, \quad \mu_q^{Sys}(\theta, \bar{\sigma}^2) = \begin{pmatrix} \mu_q^{Dif}(\theta, \bar{\sigma}^2) \\ \mu_q^{Lev}(\theta, \bar{\sigma}^2) \end{pmatrix}.$$

$$A_f^{AS}(\theta) = \begin{pmatrix} A_f^{Dif}(\theta) \\ A_f^{NL}(\theta) \vdots 0 \end{pmatrix}, \quad \mu_f^{AS}(\theta, \bar{\sigma}^2) = \begin{pmatrix} \mu_f^{Dif}(\theta, \bar{\sigma}^2) \\ \mu_f^{NL}(\theta, \bar{\sigma}^2) \end{pmatrix},$$

$$A_q^{AS}(\theta) = \begin{pmatrix} A_q^{Dif}(\theta) \\ A_q^{NL}(\theta) \vdots 0 \end{pmatrix}, \quad \mu_q^{AS}(\theta, \bar{\sigma}^2) = \begin{pmatrix} \mu_q^{Dif}(\theta, \bar{\sigma}^2) \\ \mu_q^{NL}(\theta, \bar{\sigma}^2) \end{pmatrix}.$$

T = 5. Using similar calculations, we obtain:

$$A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \end{pmatrix}, \quad \mu_f^{Dif}(\theta, \bar{\sigma}^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$A_f^{Lev}(\theta) = \begin{pmatrix} 1-\theta & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \end{pmatrix}, \quad \mu_f^{Lev}(\theta, \bar{\sigma}^2) = (1-\theta) \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \\ \sigma_4^2 \end{pmatrix},$$

$$A_f^{NL}(\theta) = \begin{pmatrix} \theta(\theta-1) & 1-\theta & 0 & 0 \\ 0 & \theta(\theta-1) & 1-\theta & 0 \end{pmatrix}, \quad \mu_f^{NL}(\theta, \bar{\sigma}^2) = (1-\theta) \begin{pmatrix} \sigma_3^2 - \theta\sigma_2^2 \\ \sigma_4^2 - \theta\sigma_3^2 \end{pmatrix}.$$

□

General T. Along the lines of the above derivations, it is also possible to construct the expressions of $A_f^j(\theta), A_q^j(\theta)$, $\mu_f^j(\theta, \bar{\sigma}^2)$, and $\mu_q^j(\theta, \bar{\sigma}^2)$ for larger values of T which we, for reasons of brevity, refrain from.

Orthogonal complements of $A_f^{AS}(\theta)$ and $A_f^{Sys}(\theta)$ for $T = 4$ and 5 and the specification of the robust sample moments. We specify the orthogonal complements as in (37), which we repeat here for convenience:

$$A_f^j(\theta)_\perp = (G_{f,T}^j(\theta) \vdots G_{2,T}^j),$$

where T indicates the number of time periods and $G_{2,T}^j$ is such that $G_{2,T}^{j\prime} \mu_f^j(\theta, \bar{\sigma}^2) = 0$. This notation is used in the proofs of subsequent theorems.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

T = 4. From the expressions of $A_f^j(\theta)$ and $\mu_f^j(\theta,\bar{\sigma}^2)$ in (36), $G_{f,T=4}^j(\theta)$ and $G_{2,T=4}^j$, for $j = AS, Sys$, result as:

$$G_{f,T=4}^{AS}(\theta) = \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad G_{2,T=4}^{AS} = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix},$$

$$G_{f,T=4}^{Sys}(\theta) = \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ -\theta \\ 1 \end{pmatrix}, \quad G_{2,T=4}^{Sys} = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

From these expressions and (36), it is easily seen that

$$A_f^{AS}(\theta)'_{\perp}\mu_f^{AS}(\theta,\bar{\sigma}^2) = \begin{pmatrix} (1-\theta)\left(\sigma_3^2 - \theta\sigma_2^2\right) \\ 0 \end{pmatrix},$$

$$A_f^{Sys}(\theta)'_{\perp}\mu_f^{Sys}(\theta,\bar{\sigma}^2) = \begin{pmatrix} \sigma_3^2 - \theta\sigma_2^2 \\ 0 \end{pmatrix},$$

from which follows that $A_f^j(\theta)'_{\perp}\mu_f^j(\theta,\bar{\sigma}^2) \neq 0$, for all $\theta \neq \theta_{0,N}, j = AS, Sys$.

T = 5. The expressions for $A_f^j(\theta)$, $\mu_f^j(\theta,\bar{\sigma}^2)$, $G_{f,T=5}^j(\theta)$, and $G_{2,T=5}^j$, for $j = AS, Sys$, are:

$$A_f^{AS}(\theta) = \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ \theta(\theta-1) & 1-\theta & 0 & 0 \\ 0 & \theta(\theta-1) & 1-\theta & 0 \end{pmatrix}, \quad \mu_f^{AS}(\theta,\bar{\sigma}^2) = (1-\theta)\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sigma_3^2 - \theta\sigma_2^2 \\ \sigma_4^2 - \theta\sigma_3^2 \end{pmatrix},$$

$$G_{f,T=5}^{AS}(\theta) = \begin{pmatrix} -(1-\theta) & 0 & 0 \\ 0 & -(1-\theta) & 0 \\ 0 & 0 & -(1-\theta) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad G_{2,T=5}^{AS} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 1-\theta & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \end{pmatrix}, \quad \mu_f^{Sys}(\theta, \bar{\sigma}^2) = (1-\theta) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ \sigma_3^2 \\ \sigma_4^2 \end{pmatrix},$$

$$G_{f,T=5}^{Sys}(\theta) = \begin{pmatrix} -(1-\theta) & 0 & 0 \\ 0 & -(1-\theta) & 0 \\ 0 & 0 & -(1-\theta) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\theta & 0 & 0 \\ 1 & -\theta & -\theta \\ 0 & 1 & 1 \end{pmatrix}, \quad G_{2,T=5}^{Sys} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Straightforward algebra shows that $A_f^j(\theta)'_\perp \mu_f^j(\theta, \bar{\sigma}^2) \neq 0$, for all $\theta \neq \theta_{0,N}$, $j = AS, Sys$.

The robust sample moments are defined as

$$g_{f,T}^j(\theta) = A_f(\theta)_\perp^{j\prime} f_N^j(\theta),$$

with $A_f(\theta)_\perp^j = (G_{f,T}^j(\theta) \vdots G_{2,T}^j)$. For the Sys moment conditions, $G_{f,T}^j(\theta)$ is a linear function of θ and $G_{2,T}^j$ does not depend on θ. Since $f_N^j(\theta)$ is linear in θ as well for the Sys sample moments, the part of $g_{f,T}^j(\theta)$ resulting from $G_{f,T}^j(\theta)' f_N^j(\theta)$ is quadratic in θ, while the part that results from $G_{2,T}^{j\prime} f_N^j(\theta)$ is linear in θ. Given the specification of $G_{f,T}^j(\theta)$, $G_{2,T}^j$, and $f_N^j(\theta)$, it is then straightforward to compute the specification of a, b, and d.

For the AS moment conditions, $G_{f,T}^j(\theta)$ is a linear function of θ, and $G_{2,T}^j$ does not depend on θ. For the AS sample moments, $f_N^j(\theta)$ is quadratic in θ, but the part of $g_{f,T}^j(\theta)$ resulting from $G_{f,T}^j(\theta)' f_N^j(\theta)$ is not of third order in θ as expected but just a quadratic function of θ. The part of $g_{f,T}^j(\theta)$ that results from $G_{2,T}^{j\prime} f_N^j(\theta)$ is linear in θ. Given the specification of $G_{f,T}^j(\theta)$, $G_{2,T}^j$, and $f_N^j(\theta)$, it is then again straightforward to compute the specification of a, b, and d.

Proof of Theorem 2. Under mean stationarity, we can write

$$\begin{aligned}
\Delta y_{i2} &= (\theta_{0,N} - 1)u_{i1} + u_{i2}, \\
\Delta y_{i3} &= \theta_{0,N}(\theta_{0,N} - 1)u_{i1} + (\theta_{0,N} - 1)u_{i2} + u_{i3}, \\
\Delta y_{i4} &= \theta_{0,N}^2(\theta_{0,N} - 1)u_{i1} + \theta_{0,N}(\theta_{0,N} - 1)u_{i2} + (\theta_{0,N} - 1)u_{i3} + u_{i4}, \\
\Delta y_{i5} &= \theta_{0,N}^3(\theta_{0,N} - 1)u_{i1} + \theta_{0,N}^2(\theta_{0,N} - 1)u_{i2} + \theta_{0,N}(\theta_{0,N} - 1)u_{i3}
\end{aligned}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$
\begin{aligned}
&\quad + (\theta_{0,N} - 1)u_{i4} + u_{i5},\\
y_{i3} - y_{i1} &= (1 + \theta_{0,N})(\theta_{0,N} - 1)u_{i1} + \theta_{0,N}u_{i2} + u_{i3},\\
y_{i4} - y_{i1} &= (1 + \theta_{0,N} + \theta_{0,N}^2)(\theta_{0,N} - 1)u_{i1} + \theta_{0,N}^2 u_{i2} + \theta_{0,N}u_{i3} + u_{i4},\\
y_{i4} - y_{i2} &= (\theta_{0,N} + \theta_{0,N}^2)(\theta_{0,N} - 1)u_{i1} + (\theta_{0,N}^2 - 1)u_{i2} + \theta_{0,N}u_{i3} + u_{i4},\\
y_{i5} - y_{i1} &= (1 + \theta_{0,N} + \theta_{0,N}^2 + \theta_{0,N}^3)(\theta_{0,N} - 1)u_{i1} + \theta_{0,N}^3 u_{i2} + \theta_{0,N}^2 u_{i3} + \theta_{0,N}u_{i4} + u_{i5},\\
y_{i5} - y_{i2} &= (\theta_{0,N} + \theta_{0,N}^2 + \theta_{0,N}^3)(\theta_{0,N} - 1)u_{i1} + (\theta_{0,N}^3 - 1)u_{i2} + \theta_{0,N}^2 u_{i3} + \theta_{0,N}u_{i4} + u_{i5}.
\end{aligned}
$$

□

The robust sample moments consist of products of the above expressions. To obtain the probability limits in Theorem 2 of the elements comprising the robust sample moments, we use that

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N} - 1)u_{it}^2 &\xrightarrow[p]{} 0,\\
\frac{1}{N}\sum_{i=1}^{N}(\theta_{0,N} - 1)u_{it}u_{is} &\xrightarrow[p]{} 0,
\end{aligned}
$$

for all s and t, $t > 1$, $t \neq s$, which is implied by Assumption 1. Therefore, the a, b, and d components of the robust sample moments simplify to:

T = 4, Sys:.

$$
\begin{aligned}
a &= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(\Delta y_{i2})^2\\ 0\end{pmatrix} = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(\theta_{0,N} - 1)^2 u_{i1}^2 + u_{i2}^2\\ 0\end{pmatrix} + O_p(N^{-1/2}),\\
b &= -\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(y_{i3} - y_{i1})^2\\ \Delta y_{i2}\Delta y_{i3}\end{pmatrix}\\
&= -\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}((1 + \theta_{0,N})^2(\theta_{0,N} - 1)^2 u_{i1}^2 + \theta_{0,N}^2 u_{i2}^2 + u_{i3}^2\\ \theta_{0,N}(\theta_{0,N} - 1)^2 u_{i1}^2 + (\theta_{0,N} - 1)u_{i2}^2\end{pmatrix} + O_p(N^{-1/2}),\\
d &= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(y_{i4} - y_{i1})\Delta y_{i3}\\ \Delta y_{i2}\Delta y_{i4}\end{pmatrix}\\
&= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\theta_{0,N}(1 + \theta_0 + \theta_{0,N}^2)(\theta_{0,N} - 1)^2 u_{i1}^2 + \theta_{0,N}^2(\theta_{0,N} - 1)u_{i2}^2 + \theta_{0,N}u_{i3}^2\\ \theta_{0,N}^2(\theta_{0,N} - 1)^2 u_{i1}^2 + \theta_{0,N}(\theta_{0,N} - 1)u_{i2}^2\end{pmatrix}\\
&\quad + O_p(N^{-1/2}),
\end{aligned}
$$

where the $O_p(N^{-1/2})$ remainder terms result from the interaction terms between the different errors, like $\frac{1}{N}\sum_{i=1}^{N} u_{i2}u_{i3}$, which converge at rate $N^{-\frac{1}{2}}$, since their correlation equals zero.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Using next that, because of Assumption 1(c), $\frac{1}{N}\sum_{i=1}^{N}(1-\theta_{0,N})^2 u_{i1}^2 \xrightarrow[p]{} 0$, and $\theta_{0,N} = 1+\frac{l}{N^\tau}$, with l a fixed constant, $l<0$, we have that

$$
\begin{aligned}
a &= \begin{pmatrix} \sigma_2^2 \\ 0 \end{pmatrix} + O_p(N^{-1/2}), \\
b &= -\begin{pmatrix} \left(1+2\frac{l}{N^\tau}+\frac{l^2}{N^{2\tau}}\right)\sigma_2^2+\sigma_3^2 \\ \frac{l}{N^\tau}\sigma_2^2 \end{pmatrix} + O_p(N^{-1/2}), \\
d &= \begin{pmatrix} \left(\frac{l}{N^\tau}+2\frac{l^2}{N^{2\tau}}+\frac{l^3}{N^{3\tau}}\right)^2\sigma_2^2+(1+\frac{l}{N^\tau})\sigma_3^2 \\ (\frac{l}{N^\tau}+\frac{l^2}{N^{2\tau}})\sigma_2^2 \end{pmatrix} + O_p(N^{-1/2}),
\end{aligned}
$$

so, if $\tau > \frac{1}{2}$,

$$
\begin{aligned}
a &= \begin{pmatrix} \sigma_2^2 \\ 0 \end{pmatrix} + O_p(N^{-1/2}), \\
b &= -\begin{pmatrix} \sigma_2^2+\sigma_3^2 \\ 0 \end{pmatrix} + O_p(N^{-1/2}), \\
d &= \begin{pmatrix} \sigma_3^2 \\ 0 \end{pmatrix} + O_p(N^{-1/2}).
\end{aligned}
$$

T = 4, AS:.

$$
\begin{aligned}
a &= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i3}-y_{i1})\Delta y_{i2} \\ 0 \end{pmatrix} \\
&= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (1+\theta_{0,N})(1-\theta_{0,N})^2 u_{i1}^2 + \theta_{0,N}u_{i2}^2 \\ 0 \end{pmatrix} + O_p(N^{-1/2}), \\
b &= -\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i3}-y_{i1})\Delta y_{i3}+(y_{i4}-y_{i1})\Delta y_{i2} \\ \Delta y_{i2}\Delta y_{i3} \end{pmatrix} \\
&= -\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (1-\theta_{0,N})^2[(1+2\theta_{0,N}(1+\theta_{0,N})]u_{i,1}^2+(2\theta_{0,N}^2-\theta_{0,N})u_{i,2}^2+u_{i,3}^2 \\ \theta_{0,N}(\theta_{0,N}-1)^2u_{i1}^2+(\theta_{0,N}-1)u_{i2}^2 \end{pmatrix} \\
&\quad + O_p(N^{-1/2}), \\
d &= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i3} \\ \Delta y_{i2}\Delta y_{i4} \end{pmatrix} \\
&= \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \theta_{0,N}(1+\theta_0+\theta_{0,N}^2)(\theta_{0,N}-1)^2u_{i1}^2+\theta_{0,N}^2(\theta_{0,N}-1)u_{i2}^2+\theta_{0,N}u_{i3}^2 \\ \theta_{0,N}^2(\theta_{0,N}-1)^2u_{i1}^2+\theta_{0,N}(\theta_{0,N}-1)u_{i2}^2 \end{pmatrix} \\
&\quad + O_p(N^{-1/2}),
\end{aligned}
$$

so also,

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$
\begin{aligned}
a = &\quad \binom{\sigma_2^2}{0} + O_p(N^{-1/2}),\\
b = &\quad -\binom{\sigma_2^2+\sigma_3^2}{0} + O_p(N^{-1/2}),\\
d = &\quad \binom{\sigma_3^2}{0} + O_p(N^{-1/2}).
\end{aligned}
$$

We use similar calculations for $T = 5$ to obtain that:

T = 5, Sys:.

$$
a = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (\Delta y_{i2})^2 \\ (y_{i3}-y_{i1})\Delta y_{i3} \\ (\Delta y_{i3})^2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \\ \sigma_3^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),
$$

$$
b = -\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i3}-y_{i1})^2 \\ (y_{i4}-y_{i1})(y_{i4}-y_{i2}) \\ (y_{i4}-y_{i2})^2 \\ \Delta y_{i2}\Delta y_{i4} \\ \Delta y_{i3}\Delta y_{i4} \end{pmatrix} = -\begin{pmatrix} \sigma_2^2+\sigma_3^2 \\ \sigma_3^2+\sigma_4^2 \\ \sigma_3^2+\sigma_4^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),
$$

$$
d = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i3} \\ (y_{i5}-y_{i1})\Delta y_{i4} \\ (y_{i5}-y_{i2})\Delta y_{i4} \\ \Delta y_{i2}\Delta y_{i5} \\ \Delta y_{i3}\Delta y_{i5} \end{pmatrix} = \begin{pmatrix} \sigma_3^2 \\ \sigma_4^2 \\ \sigma_4^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}).
$$

T = 5, AS:.

$$
a = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i3}-y_{i1})\Delta y_{i2} \\ (y_{i4}-y_{i1})\Delta y_{i3} \\ (y_{i4}-y_{i2})\Delta y_{i3} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \\ \sigma_3^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),
$$

$$
b = -\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i2}+(y_{i3}-y_{i1})\Delta y_{i3} \\ (y_{i4}-y_{i1})\Delta y_{i4}+(y_{i5}-y_{i1})\Delta y_{i3} \\ (y_{i4}-y_{i2})\Delta y_{i4}+(y_{i5}-y_{i2})\Delta y_{i3} \\ \Delta y_{i2}\Delta y_{i4} \\ \Delta y_{i3}\Delta y_{i4} \end{pmatrix} = -\begin{pmatrix} \sigma_2^2+\sigma_3^2 \\ \sigma_3^2+\sigma_4^2 \\ \sigma_3^2+\sigma_4^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}),
$$

$$
d = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} (y_{i4}-y_{i1})\Delta y_{i3} \\ (y_{i5}-y_{i1})\Delta y_{i4} \\ (y_{i5}-y_{i2})\Delta y_{i4} \\ \Delta y_{i2}\Delta y_{i5} \\ \Delta y_{i3}\Delta y_{i5} \end{pmatrix} = \begin{pmatrix} \sigma_3^2 \\ \sigma_4^2 \\ \sigma_4^2 \\ 0 \\ 0 \end{pmatrix} + O_p(N^{-\frac{1}{2}}).
$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Proof of Theorem 3. The proof of Theorem 3 establishes the probability limits of a, b, and d, for $\theta_{0,N} = 1 + \frac{l}{N^\tau}$, $l < 0$, and $\tau > \frac{1}{2}$. Denoting these probability limits by, a_p, b_p, and d_p, the large sample behavior of a, b, and d is characterized by, for $\theta_{0,N} = 1 + \frac{l}{N^\tau}$ with $\tau > \frac{1}{2}$:

$$\sqrt{N}(a - a_p) \underset{d}{\to} \varepsilon_a, \sqrt{N}(b - b_p) \underset{d}{\to} \varepsilon_b, \sqrt{N}(d - d_p) \underset{d}{\to} \varepsilon_d,$$

with $(\varepsilon_a, \varepsilon_b, \varepsilon_d)$ jointly normal, mean-zero random variables, which follows straightforwardly from an appropriate central limit theorem applied to the highest-order remainder terms in the proof of Theorem 2, which are all sample averages over i.i.d. mean-zero random variables. We want to determine the appropriate rate for ξ in $g_{f,T}(\theta(e))$, so we can analyze its behavior in a neighborhood of the true value $\theta_{0,N} = 1 + \frac{l}{N^\tau}$, $l < 0$, with $\tau > \frac{1}{2}$ while N goes to infinity, with

$$\theta(e) = 1 + \tfrac{e}{N^\xi}.$$

Substituting $\theta(e)$ and the above large sample characterizations of a, b, and d in (38), we can write:

$$g_{f,T}(\theta(e)) = (1 + \tfrac{e}{N^\xi})^2(a_p + \tfrac{\varepsilon_a}{\sqrt{N}}) + (1 + \tfrac{e}{N^\xi})(b_p + \tfrac{\varepsilon_b}{\sqrt{N}}) + d_p + \tfrac{\varepsilon_d}{\sqrt{N}} + o_p(N^{-1/2}).$$

To determine ξ, we impose two conditions: (1) $\sqrt{N} g_{f,T}(\theta(e))$ converges to a nondegenerate bounded random variable of order $O_p(1)$; and (2) $g_{f,T}(\theta(e))$ is informative about the value of e when N gets large. We discriminate between two different cases for σ_t^2 :

1. For $\sigma_t^2 = \sigma^2$, $t = 2, \ldots, T$:

$$\begin{aligned} & g_{f,T}(\theta(e)) \\ & = (1 + \frac{e}{N^\xi})^2(a_p + \frac{\varepsilon_a}{\sqrt{N}}) + (1 + \frac{e}{N^\xi})(b_p + \frac{\varepsilon_b}{\sqrt{N}}) + d_p + \frac{\varepsilon_d}{\sqrt{N}} + o_p(N^{-1/2}) \\ & = a_p + b_p + d_p + \tfrac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d) + \Big(\frac{e}{N^\xi}\Big)^2 a_p \\ & \quad + \frac{e}{N^\xi}(b_p + 2a_p) + \frac{e}{N^\xi \sqrt{N}}(\varepsilon_b + 2\varepsilon_a) + \frac{e^2}{N^{2\xi} N^{1/2}} \varepsilon_a + o_p(N^{-1/2}), \end{aligned}$$

since $a_p + b_p + d_p = 0$ and $b_p + 2a_p = 0$, we distinguish three settings:

$\xi < 1/4$:

$$g_{f,T}(\theta(e)) = \frac{e^2}{N^{2\xi}} a_p + o_p(N^{-2\xi}),$$

$\xi = 1/4$:

$$\begin{aligned} g_{f,T}(\theta(e)) &= \tfrac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d + e^2 a_p) + \frac{e}{\sqrt{N}\sqrt[4]{N}}(\varepsilon_b + 2\varepsilon_a) + \frac{e^2 \varepsilon_a}{N} + o_p(N^{-1/2}) \\ &= \tfrac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d + e^2 a_p) + o_p(N^{-1/2}), \end{aligned}$$

$\xi > 1/4$:

$$g_{f,T}(\theta(e)) = \tfrac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d) + o_p(N^{-1/2}).$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

This shows that the appropriate rate corresponds with $\xi = 1/4$. For a smaller value of ξ, $\sqrt{N}g_{f,T}(\theta(e))$ diverges. For a larger value, $\sqrt{N}g_{f,T}(\theta(e))$ converges to a mean-zero normal random variable unaffected by the choice of e. Although, in this case, $\sqrt{N}g_{f,T}(\theta(e))$ is not informative about e, we do not need to worry about e, because standard asymptotics apply.

2. When $\sigma_t^2 \neq \sigma_s^2$, for at least one $t \neq s$, $a_p + b_p + d_p = 0$ but $b_p + 2a_p \neq 0$, we can establish along the lines of the above that the appropriate rate corresponds with $\xi = 1/2$:

$$
\begin{aligned}
&g_{f,T}(\theta(e)) \\
&\quad = (1+\frac{e}{\sqrt{N}})^2(a_p+\frac{\varepsilon_a}{\sqrt{N}})+(1+\frac{e}{\sqrt{N}})(b_p+\frac{\varepsilon_b}{\sqrt{N}})+d_p+\frac{\varepsilon_d}{\sqrt{N}}+o_p(N^{-1/2}) \\
&\quad = a_p+b_p+d_p+\tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e(b_p+a_p)) \\
&\qquad +\frac{e}{N}(2\varepsilon_a+\varepsilon_b+eE(a))+\frac{e^2\varepsilon_a}{N\sqrt{N}}+o_p(N^{-1/2}) \\
&\quad = \tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e(b_p+2a_p))+\frac{e}{N}(2\varepsilon_a+\varepsilon_b+ea_p)+\frac{e^2\varepsilon_a}{N\sqrt{N}}+o_p(N^{-1/2}) \\
&\quad = \tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e(b_p+2a_p))+o_p(N^{-1/2}).
\end{aligned}
$$

□

Proof of Theorem 4. Denote with $g_{f,T}(\theta(e))$ the moments in (38) evaluated at $\theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$. When $\sigma_t^2 = \sigma^2$ and substituting the large sample characterization of a, b, and d, $\sqrt{N}g_{f,T}(\theta(e))$ can be expressed as:

$$\sqrt{N}g_{f,T}(\theta(e)) = e^2a_p + \varepsilon_a(1+\frac{2e}{\sqrt[4]{N}}+\frac{e^2}{\sqrt{N}})+\varepsilon_b(1+\frac{e}{\sqrt[4]{N}})+\varepsilon_d+o_p(1).$$

Define

$$\phi(N) = e^2a_p+\varepsilon_a\left(1+\frac{2e}{\sqrt[4]{N}}+\frac{e^2}{\sqrt{N}}\right)+\varepsilon_b\left(1+\frac{e}{\sqrt[4]{N}}\right)+\varepsilon_d.$$

Since $(\varepsilon_a, \varepsilon_b, \varepsilon_d)$ are jointly normal distributed,

$$\phi(N) \sim N(e^2a_p,\ B(N)'V_{abd}B(N))$$

with

$$B(N) = (\iota_3 \otimes I_{p_{\max}}) + \frac{e}{\sqrt[4]{N}}\left[(2+\tfrac{e}{\sqrt[4]{N}})(e_{1,3}\otimes I_{p_{\max}})+(e_{2,3}\otimes I_{p_{\max}})\right],$$

and V_{abd} the covariance matrix of $(\varepsilon_a' \vdots \varepsilon_b' \vdots \varepsilon_d')'$, ι_3 a 3×1 dimensional vector of ones, $I_{p_{\max}}$ the $p_{\max}\times p_{\max}$ dimensional identity matrix, $p_{\max}$ equals the number of elements of a, and $e_{1,3}$ and $e_{2,3}$ the first and second 3×1 dimensional unity vectors.

Hence,

$$\sqrt{N}g_{f,T}(\theta(e)) = \phi(N)+o_p(1),$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

so in a sample of size N, $\sqrt{N}g_{f,T}(\theta(e))$ is normally distributed up to an $o_p(1)$ term. While some of the components in $\phi(N)$ are essentially also $o_p(1)$, it is important to incorporate them for an accurate approximation of the distribution of $\sqrt{N}g_{f,T}(\theta(e))$ for a given sample of size N, since the low-order components, of order $N^{-1/4}$, converge very slowly to zero.

The individual moments $g_{f,n}(\theta(e))$ in the sample average $g_{f,T}(\theta(e)) = \frac{1}{N}\sum_{n=1}^{N} g_{f,n}(\theta(e))$ can be specified as:

$$\begin{aligned}
g_{f,n}(\theta(e)) &= (1+\tfrac{e}{\sqrt[4]{N}})^2 a_n + (1+\tfrac{e}{\sqrt[4]{N}})b_n + d_n \\
&= (1+\tfrac{e}{\sqrt[4]{N}})^2[a_p+\varepsilon_{a_n}] + (1+\tfrac{e}{\sqrt[4]{N}})[b_p+\varepsilon_{b_n}] + [d_p+\varepsilon_{d_n}] \\
&= (a_p+b_p+d_p) + \tfrac{e}{\sqrt[4]{N}}(2a_p+b_p) + \tfrac{e^2}{\sqrt{N}}a_p \\
&\quad + \varepsilon_{a_n}+\varepsilon_{b_n}+\varepsilon_{d_n} + \tfrac{e}{\sqrt[4]{N}}(2\varepsilon_{a_n}+\varepsilon_{b_n}) + \tfrac{e^2}{\sqrt{N}}\varepsilon_{a_n} \\
&= \tfrac{e^2}{\sqrt{N}}a_p + \varepsilon_{a_n}+\varepsilon_{b_n}+\varepsilon_{d_n} + \tfrac{e}{\sqrt[4]{N}}(2\varepsilon_{a_n}+\varepsilon_{b_n}) + \tfrac{e^2}{\sqrt{N}}\varepsilon_{a_n},
\end{aligned}$$

with $a = \frac{1}{N}\sum_{n=1}^{N} a_n$, $b = \frac{1}{N}\sum_{n=1}^{N} b_n$, $d = \frac{1}{N}\sum_{n=1}^{N} d_n$, $\varepsilon_{a_n} = a_n - a_p$, $\varepsilon_{b_n} = b_n - b_p$, and $\varepsilon_{d_n} = d_n - d_p$, so taking $g_{f,n}(\theta(e))$ in deviation from its sample average $g_{f,T}(\theta(e))$ results in

$$\begin{aligned}
g_{f,n}(\theta(e)) - g_{f,T}(\theta(e)) &= \varepsilon_{a_n} - \varepsilon_a + \varepsilon_{b_n} - \varepsilon_b + \varepsilon_{d_n} - \varepsilon_d \\
&\quad + \tfrac{e}{\sqrt[4]{N}}(2(\varepsilon_{a_n}-\varepsilon_a) + \varepsilon_{b_n} - \varepsilon_b) + \tfrac{e^2}{\sqrt{N}}(\varepsilon_{a_n}-\varepsilon_a) + o_p(N^{-1/2}).
\end{aligned}$$

From the above, it then straightforwardly follows that

$$\begin{aligned}
\hat{V}_{gg}(e) &= \tfrac{1}{N}\sum_{i=1}^{N}\left(g_{f,n}(\theta(e)) - g_{f,T}(\theta(e))\right)\left(g_{f,n}(\theta(e)) - g_{f,T}(\theta(e))\right)' \\
&= B(N)' V_{abd} B(N) + o_p(1),
\end{aligned}$$

so the distribution of the GMM-AR statistic testing H_p for a sample of size N is characterized by

$$\chi^2(\delta(N), p_{\max}) + o_p(1),$$

with $\delta(N) = e^4 a_p' \left[B(N)' V_{abd} B(N)\right]^{-1} a_p$. □

Proof of Theorem 5. When we instead of the full vector $g_{f,T}(\theta(e))$ use a linear combination of it, say $w' g_{f,T}(\theta(e))$ with w an orthonormal $p_{\max} \times 1$ vector, the approximating distribution of the GMM-AR statistic for testing $\mathrm{H}_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ that uses $w' g_{f,T}(\theta(e))$ as the moment vector reads

$$\chi^2(e^4 (w'a_p)' \left[w' B(N)' V_{abd} B(N) w\right]^{-1} (w'a_p), 1).$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

The optimal combination w is the one that leads to the largest value of the noncentrality parameter. The noncentrality parameter can be specified as

$$e^4(w'a_p)'\left[w'B(N)'V_{abd}B(N)w\right]^{-1}(w'a_p) \quad = e^4\frac{(w'a_p)^2}{w'B(N)'V_{abd}B(N)w}.$$

The maximal value of $\frac{(w'a_p)^2}{w'B(N)'V_{abd}B(N)w}$ results from the largest root of the generalized eigenvalue problem

$$\left|\lambda B(N)'V_{abd}B(N) - a_p a_p'\right| = 0,$$

and the optimal value of w equals the eigenvector associated with the largest root. Since a_p is only a vector, just one root of the generalized eigenvalue problem is nonzero, so it is also the largest one. This root results from using

$$w = (B(N)'V_{abd}B(N))^{-1}a_p,$$

and the largest root then equals

$$\lambda_{\max} = a_p'(B(N)'V_{abd}B(N))^{-1}a_p,$$

so the maximal value of the noncentrality parameter is

$$\delta(N) = e^4 a_p'(B(N)'V_{abd}B(N))^{-1}a_p = (e\sigma)^4\begin{pmatrix}\iota_p\\0\end{pmatrix}'(B(N)'V_{abd}B(N))^{-1}\begin{pmatrix}\iota_p\\0\end{pmatrix},$$

since $a_p = \sigma^2\binom{\iota_p}{0}$ with ι_p a $p\times 1$ dimensional vector of ones and p the number of columns of $G_{f,T}(\theta)$. □

Proof of Theorem 6. Before we start out to prove Theorem 6, we first state an addendum to Theorem 1, which incorporates some higher-order components of order $O_p(N^{-1/2})$ that are needed for some of the intermediate results. □

Addendum to Theorem 1: Theorem 1* (Representation theorem). *Under Assumptions 1 and 2(a), we can characterize the large sample behavior of the Dif, Lev, NL, AS, and Sys sample moments and their derivatives by:*

$$\begin{pmatrix} f_N^j(\theta)\\ q_N^j(\theta)\end{pmatrix} = \begin{pmatrix} A_f^j(\theta)\\ A_q^j(\theta)\end{pmatrix}\left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}(\psi - h_N(\theta_{0,N})\sigma_{1,n}\iota_{T-1}\psi_c) + \iota_{T-1}d_2\right]$$
$$+\begin{pmatrix}\mu_f^j(\theta,\bar\sigma^2)\\ \mu_q^j(\theta,\bar\sigma^2)\end{pmatrix} + \frac{1}{\sqrt{N}}\begin{pmatrix} B_f^j(\theta)\\ B_q^j(\theta)\end{pmatrix}\psi_{uu} + o_p(N^{-1/2}),$$

with $j = Dif, Lev, NL, AS, Sys$, and $B_f^j(\theta), B_q^j(\theta)$: $k_j\times m_j$ and $k_j\times m_j$, $k_j\times 1$ dimensional matrices, and ψ_{uu} is a mean-zero, finite variance, normal random vector that is possibly dependent on ψ.

Proof of large sample distribution KLM statistic. For the construction of the large sample distribution of the KLM statistic under Assumptions 1 and 2(a), we use that the part of the sample moments spanned by $A_f^j(\theta(e))$ and the part spanned by $A_f^j(\theta(e))_\perp$ converge at different rates. We use the normalized large sample behavior of each of these parts to construct it. This amounts to premultiplying the sample moments in the expression of the

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

KLM statistic by $(A_f^j(\theta(e)) \vdots A_f^j(\theta(e))_\perp)$ to which it is invariant if $(A_f^j(\theta(e)) \vdots A_f^j(\theta(e))_\perp)$ is invertible. The specification of $A_f^j(\theta(e))_\perp$ as equal to $(G_{f,T}^j(\theta(e)) \vdots G_{2,T}^j)$ (see (37)) is such that $(A_f^j(\theta(e)) \vdots A_f^j(\theta(e))_\perp)$ is invertible for the Sys moment conditions but not for the AS moment conditions both when $T = 4$ and 5, since $A_f^j(\theta(e))$ does not have full column rank. To have an invertible specification of $(A_f^j(\theta(e)) \vdots A_f^j(\theta(e))_\perp)$, we use that we can specify $A_f^j(\theta(e))$ for the AS moments as:

$$\mathbf{T=4}: \quad A_f^{AS}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \\ \theta(\theta-1) & 1-\theta & 0 \end{pmatrix} = A_{f,T=4}^{AS}(\theta)_1 A_{f,T=4}^{AS}(\theta)_2,$$

$$T=5: \quad A_f^{AS}(\theta) = \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ \theta(\theta-1) & 1-\theta & 0 & 0 \\ 0 & \theta(\theta-1) & 1-\theta & 0 \end{pmatrix} = A_{f,T=5}^{AS}(\theta)_1 A_{f,T=5}^{AS}(\theta)_2,$$

where

$$T=4: \quad A_{f,T=4}^{AS}(\theta)_1 = \begin{pmatrix} -\theta & 0 \\ 0 & 1 \\ 0 & 1 \\ \theta(\theta-1) & 0 \end{pmatrix},$$

$$A_{f,T=4}^{AS}(\theta)_2, \begin{pmatrix} 1 & -\theta^{-1} & 0 \\ 0 & -\theta & 1 \end{pmatrix},$$

$$T=5: \quad A_{f,T=5}^{AS}(\theta)_1 = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 0 \\ 0 & -\theta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ \theta(\theta-1) & 1-\theta & 0 \\ 0 & \theta(\theta-1) & 0 \end{pmatrix}$$

$$A_{f,T=5}^{AS}(\theta)_2 = \begin{pmatrix} 1 & 0 & -\theta^{-2} & 0 \\ 0 & 1 & -\theta^{-1} & 0 \\ 0 & 0 & -\theta & 1 \end{pmatrix},$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

so unlike $A_f^{AS}(\theta)$, $A_f^{AS}(\theta)_1$ has full column rank. For the Sys moments, for which $A_f^{Sys}(\theta)$ has full column rank, we use $A_f^{Sys}(\theta)_1 = A_f^{Sys}(\theta)$. The matrix $(A_f^j(\theta(e))_1 \vdots A_f^j(\theta(e))_\perp)$ is now invertible for both $j = AS, Sys$, so we use it to construct the large sample behavior of the KLM statistic to test $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ while the true value of θ is drifting to one in line with Assumption 2(a). We separately construct the behavior of the following four components:

1. $\sqrt{N}\hat{V}_{ff}(\theta(e))^{-1} f_N(\theta(e))$,
2. $q_N(\theta(e))$,
3. $\hat{V}_{\theta f}(\theta(e))$,
4. $\hat{D}_N(\theta(e))$,

which provide the building blocks for the large sample distribution of the KLM statistic. For each of these components, we determine their limit behavior when multiplied by $(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(e)_\perp)$ for the last three components and its inverse for the first one. Taken all together, this implies that $(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(e)_\perp)$ cancels out of the overall expression of the KLM statistic.

1. To determine the limit behavior of $\sqrt{N}\hat{V}_{ff}(\theta(e))^{-1} f_N(\theta(e))$, we disentangle the components with different convergence rates which we do by premultiplying it by $(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(e)_\perp)^{-1}$:

$$
\begin{aligned}
&(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(e)_\perp)^{-1}\sqrt{N}\hat{V}_{ff}(\theta(e))^{-1} f_N(\theta(e)) \\
&\quad = \left[(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e)_1 \vdots A_f(e)_\perp)\right]^{-1} \\
&\qquad \times \left[\sqrt{N}(h_N(\theta_{0,N})A_f(e)_1 \vdots A_f(e)_\perp)' f_N(e)\right].
\end{aligned}
$$

We next determine the large sample behavior of the different components under Assumptions 1 and 2(a). Our specification of $A_f(\theta(e))_\perp$ is such that:

$$\sqrt{N}A_f(\theta(e))'_\perp f_N(\theta(e)) = \sqrt{N}g_{f,T}(\theta(e)),$$

so using the large sample behavior of $\sqrt{N}g_{f,T}(\theta(e))$ stated in the proof of Theorem 4, we have that the large sample behavior of $\sqrt{N}A_f(\theta(e))'_\perp f_N(\theta(e))$ for a (large) sample of size N results as:

$$\sqrt{N}A_f(\theta(e))'_\perp f_N(\theta(e)) = \left[e^2\sigma^2\binom{\iota_p}{0} + B(N)'\begin{pmatrix}\varepsilon_a\\ \varepsilon_b\\ \varepsilon_d\end{pmatrix}\right] + o_p(1).$$

The large sample behavior of $\sqrt{N}h_N(\theta_{0,N})A_f(\theta(e))'_1 f_N(\theta(e))$ results from Theorem 1 (the representation theorem) and accords with, since by Assumption 2(a) $\sqrt{N}h_N(\theta_{0,N}) \to 0$,

$$\sqrt{N}h_N(\theta_{0,N})A_f(\theta(e))'_1 f_N(\theta(e)) = \quad A_f(\theta(e))'_1 A_f(\theta(e))\bar{\psi} + o_p(1),$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

where $\bar{\psi} = \psi - h_N(\theta_{0,N})\sigma_{1,n}\iota_{T-1}\psi_c$, so upon combining:

$$\left[\sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)' f_N(\theta(e))\right]$$
$$= \left[\begin{array}{c} A_f(\theta(e))_1' A_f(\theta(e))\bar{\psi} \\ e^2\sigma^2\binom{\iota_p}{0} + B(N)'\begin{pmatrix}\varepsilon_a\\ \varepsilon_b\\ \varepsilon_d\end{pmatrix}\end{array}\right] + o_p(1).$$

We next focus on the components of $[(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})$ $A_f(e)_1 \vdots A_f(e)_\perp)]$. Since $g_{f,T}(\theta(e))$ does not depend on the initial observations y_{i1}, the (normalized) covariance of $A_f(\theta(e))_1' f_N(\theta(e))$ and $A_f(\theta(e))_\perp' f_N(\theta(e))$ equals zero:

$$h_N(\theta_{0,N})A_f(\theta(e))_1'\hat{V}_{ff}(\theta(e))A_f(\theta(e))_\perp = o_p(1).$$

Under Assumption 2(a), also:

$$h_N(\theta_{0,N})^2 A_f(\theta(e))_1'\hat{V}_{ff}(\theta(e))A_f(\theta(e))_1 = A_f(\theta(e))_1' A_f(\theta(e))\Lambda A_f(\theta(e))' A_f(\theta(e))_1 + o_p(1),$$
$$A_f(\theta(e))_\perp'\hat{V}_{ff}(e)A_f(\theta(e))_\perp = B(N)'V_{abd}B(N) + o_p(1),$$

where

$$\Lambda = var\left(\lim_{N\to\infty}\bar{\psi}\right)$$
$$= diag(\bar{\sigma}^2) + \left[\lim_{N\to\infty}\left(h_N(\theta_{0,N})^2\sigma_{1,n}^2\right)\right]\iota_{T-1}\iota_{T-1}' var(c_i),$$

so

$$(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)$$
$$= \begin{pmatrix} A_f(\theta(e))_1' A_f(\theta(e))\Lambda A_f(\theta(e))' A_f(\theta(e))_1 & 0 \\ 0 & B(N)'V_{abd}B(N)\end{pmatrix} + o_p(1).$$

Because $h_N(\theta_{0,N})A_f(\theta(e))_1' f_N(\theta(e))$ and $A_f(\theta(e))_\perp' f_N(\theta(e))$ are uncorrelated under Assumption 2(a),

$$\left[(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)\right]$$

convergences to a block diagonal matrix, so we obtain the large sample behavior of $\sqrt{N}((h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp))^{-1}$ $(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)' f_N(\theta(e))$:

$$\sqrt{N}((h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp))^{-1}$$
$$\times (h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)' f_N(\theta(e))$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$= \begin{pmatrix} \left[A_f(\theta(e))'_1 A_f(\theta(e)) \Lambda A_f(\theta(e))' A_f(\theta(e))_1\right]^{-1} A_f(\theta(e))'_1 A_f(\theta(e)) \bar{\psi} \\ \left(B(N)' V_{abd} B(N)\right)^{-1} \left(e^2\sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{pmatrix} + o_p(1).$$

2. To obtain the large sample behavior of $q_N(\theta(e))$ under Assumptions 1 and 2(a), we characterize the behavior of the different components of

$$(h_N(\theta_{0,N}) A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)' q_N(\theta(e)),$$

for which we use the representation of $q_N(\theta(e))$ in Theorem 1 (and Theorem 1*).

Under DGPs according with Assumptions 1 and 2(a), $\sqrt{N} h_N(\theta_{0,N}) A_f(\theta(e))'_1 q_N(\theta(e))$ is characterized by

$$\begin{aligned} &\sqrt{N} h_N(\theta_{0,N}) A_f(\theta(e))'_1 q_N(\theta(e)) \\ &\quad = A_f(\theta(e))'_1 \left[A_q(\theta(e)) \bar{\psi} + h_N(\theta_{0,N}) \sqrt{N} (\mu_q(\theta(e), \bar{\sigma}^2) \right. \\ &\qquad \left. - A_q(\theta(e)) \iota_{T-1} d_2 \right] + o_p(1), \end{aligned}$$

which converges to

$$A_f(\theta(e))'_1 A_q(\theta(e)) \bar{\psi},$$

since under Assumption 2(a),

$$\sqrt{N} h_N(\theta_{0,N}) (\mu_q(\theta(e), \bar{\sigma}^2) - A_q(\theta(e))) \iota_{T-1} d_2 \to 0,$$

which results from Assumption 1(b) and $h_N(\theta_{0,N}) \sqrt{N} \to 0$.

Regarding $A_f(\theta(e))'_\perp q_N(\theta(e))$, we distinguish between the AS and Sys moment conditions. For the Sys moment conditions,

$$\begin{aligned} &\sqrt[4]{N} A_f(\theta(e))'_\perp q_N(\theta(e)) \\ &\quad = \sqrt[4]{N} \begin{pmatrix} G_{f,T}(\theta(e))' q_N(\theta(e)) \\ G'_{2,T} q_N(\theta(e)) \end{pmatrix} \\ &\quad = -\begin{pmatrix} e\sigma^2 \iota_p \\ 0 \end{pmatrix} + \frac{1}{\sqrt[4]{N}} \begin{pmatrix} \frac{1}{h_N(\theta_{0,N})} G_f(\theta(e))' A_q(\theta(e)) \bar{\psi} + \varepsilon_{aq} \\ \varepsilon_{bq} \end{pmatrix} + o_p(N^{-1/4}), \end{aligned}$$

for which we used the representation for $q_N(\theta(e))$ that results from Theorem 1* in the Appendix which includes $B_q(\theta)\psi_{uu}$, since, for the Sys moment conditions, $G'_{2,T} A_q(\theta(e)) = 0$, $G'_{2,T} \mu(\theta(e), \bar{\sigma}^2) = 0$, $G_{f,T}(\theta(e))' A_q(\theta(e)) \iota_{T-1} = 0$, $G_{f,T}(\theta(e))' \mu(\theta(e), \bar{\sigma}^2) = -\frac{e}{\sqrt[4]{N}} \sigma^2 \iota_p$, and $\varepsilon_{aq} = G_f(\theta(e))' B_q(\theta(e)) \psi_{uu}$ and $\varepsilon_{bq} = G'_{2,T} B_q(\theta(e)) \psi_{uu}$ are mean-zero normal random variables that capture the remaining random parts.

For the AS moment conditions,

$$\begin{aligned} &\sqrt[4]{N} A_f(\theta(e))'_\perp q_N(\theta(e)) \\ &\quad = -\begin{pmatrix} e(2\sigma^2 - d_2) \iota_p \\ 0 \end{pmatrix} + \frac{1}{\sqrt[4]{N}} \begin{pmatrix} \frac{1}{h_N(\theta_{0,N})} G_f(\theta(e))' A_q(\theta(e)) \bar{\psi} + \varepsilon_{aq} \\ \varepsilon_{bq} \end{pmatrix} + o_p(N^{-1/4}), \end{aligned}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

since, for the AS moment conditions, $G'_{2,T}A_q(\theta(e)) = 0$, $G'_{2,T}\mu(\theta(e), \bar{\sigma}^2) = 0$, $G_{f,T}(\theta(e))'$ $A_q(\theta(e))\iota_{T-1} = \frac{e}{\sqrt[4]{N}}\iota_p$, $G_f(\theta(e))'\mu(\theta(e), \bar{\sigma}^2) = -\frac{2e}{\sqrt[4]{N}}\sigma^2\iota_p$, and $\varepsilon_{aq} = G_f(\theta(e))'$ $B_q(\theta(e))\psi_{cu}$ and $\varepsilon_{bq} = G'_{2,T}B_q(\theta(e))\psi_{cu}$ are mean-zero normal random variables that capture the remaining random parts.

Overall, the large sample behavior of $A_f(\theta(e))'_\perp q_N(\theta(e))$ for both the AS and Sys moment conditions reads:

$$\sqrt[4]{N}A_f(\theta(e))'_\perp q_N(\theta(e)) = \left[-\begin{pmatrix} \bar{e}\iota_p \\ 0 \end{pmatrix} + \frac{1}{\sqrt[4]{N}}\begin{pmatrix} \frac{1}{h_N(\theta_{0,N})}G_f(\theta(e))'A_q(\theta(e))\bar{\psi} + \varepsilon_{aq} \\ \varepsilon_{bq} \end{pmatrix}\right] + o_p(N^{-1/4}),$$

where for

$$\begin{aligned} \textbf{Sys}: \quad & \bar{e} = e\sigma^2, \\ \textbf{AS}: \quad & = e\left[2\sigma^2 - d_2\right]. \end{aligned}$$

Combining our results for the two components,

$$\begin{aligned} &(\sqrt{N}h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots \sqrt[4]{N}A_f(\theta(e))_\perp)'q_N(\theta(e)) \\ &\quad = (\sqrt{N}h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots \sqrt[4]{N}(G_{f,T}(\theta(e)) \vdots G_{2,T}))'q_N(\theta(e)) \\ &\quad = \begin{pmatrix} A_f(\theta(e))'_1A_q \\ \frac{1}{h_N(\theta_{0,N})\sqrt[4]{N}}G_f(\theta(e))'A_q \\ 0 \end{pmatrix}\bar{\psi} + \begin{pmatrix} 0 \\ -\bar{e}\iota_p + \frac{1}{\sqrt[4]{N}}\varepsilon_{aq} \\ \frac{1}{\sqrt[4]{N}}\varepsilon_{bq} \end{pmatrix} + o_p(N^{-1/4}), \end{aligned}$$

where it is again important to incorporate the higher-order components. We can also specify the above convergence as:

$$\begin{aligned} &\sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'q_N(\theta(e)) \\ &= \begin{pmatrix} A_f(\theta(e))'_1A_q \\ \frac{1}{h_N(\theta_{0,N})}G_f(\theta(e))'A_q \\ 0 \end{pmatrix}\bar{\psi} + \begin{pmatrix} 0 \\ -\sqrt[4]{N}\bar{e}\iota_p + \varepsilon_{aq} \\ \varepsilon_{bq} \end{pmatrix} + o_p(1). \end{aligned}$$

3. We next determine the behavior of $\hat{V}_{\theta f}(\theta(e))$:

$$\begin{aligned} &(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{\theta f}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp) \\ &\quad = \begin{pmatrix} A_f(\theta(e))'_1A_q \\ \frac{1}{h_N(\theta_{0,N})}G_f(\theta(e))'A_q \\ 0 \end{pmatrix}\Lambda\begin{pmatrix} A_f(\theta(e))'_1A_f(\theta(e)) \\ 0 \\ 0 \end{pmatrix}' \\ &\qquad + \begin{pmatrix} 0 & 0 \\ 0 & \begin{pmatrix} V_{aq,abd}B(N) \\ V_{bq,abd}B(N) \end{pmatrix} \end{pmatrix} + o_p(1), \end{aligned}$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

with $V_{aq,abd}$, $V_{aq,abd}$ the covariances between ε_{aq} and $(\varepsilon_a' \vdots \varepsilon_b' \vdots \varepsilon_d')'$ and ε_{bq} and $(\varepsilon_a' \vdots \varepsilon_b' \vdots \varepsilon_d')'$, respectively, which results directly from the specifications in Theorem 1 (and 1* in the Appendix) and those above.

Combining with the large sample behavior of $\sqrt{N}((h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'$ $\hat{V}_{ff}(\theta(e))\ (h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp))^{-1}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'f_N(\theta(e))$, we have:

$$
\begin{aligned}
&\sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{\theta f}(\theta(e))\hat{V}_{ff}(\theta(e))^{-1}f_N(\theta(e)) \\
&= \sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{\theta f}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp) \\
&\quad \times ((h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp))^{-1} \\
&\quad \times (h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'f_N(\theta(e)) \\
&= \begin{pmatrix} A_f(\theta(e))_1'A_q \\ \left(\frac{1}{h_N(\theta_{0,N})}G_f(\theta(e))'A_q\right) \\ 0 \end{pmatrix}\bar{\psi} + \begin{pmatrix} 0 \\ \left(V_{aq,abd}B(N)\right) \\ V_{bq,abd}B(N) \end{pmatrix} \\
&\quad \times \left(B(N)'V_{abd}B(N)\right)^{-1}\left(e^2\sigma^2\begin{pmatrix}\iota_p \\ 0\end{pmatrix} + B(N)'\begin{pmatrix}\varepsilon_a \\ \varepsilon_b \\ \varepsilon_d\end{pmatrix}\right) + o_p(1).
\end{aligned}
$$

4. For the large sample behavior of $\hat{D}_N(\theta(e))$, we next combine the behaviors of $\sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'q_N(\theta(e))$ constructed under **2** and $\sqrt{N}(h_N(\theta_{0,N})$ $A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{\theta f}(\theta(e))\hat{V}_{ff}(\theta(e))^{-1}f_N(\theta(e))$, which is constructed under **3.**Upon combining them, the large sample behavior of $\sqrt[4]{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{D}_N$ $(\theta(e))$ results as

$$
\begin{aligned}
&\sqrt[4]{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{D}_N(\theta(e)) \\
&= \frac{1}{\sqrt[4]{N}}\left\{\left[\sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'q_N(\theta(e))\right.\right. \\
&\quad \left.\left. -\sqrt{N}(h_N(\theta_{0,N})A_f(\theta(e))_1 \vdots A_f(\theta(e))_\perp)'\hat{V}_{\theta f}(\theta(e))\hat{V}_{ff}(\theta(e))^{-1}f_N(\theta(e))\right]\right. \\
&= \frac{1}{\sqrt[4]{N}}\left\{\begin{pmatrix} A_f(\theta(e))_1'A_q \\ \left(\frac{1}{h_N(\theta_{0,N})}G_f(\theta(e))'A_q\right) \\ 0 \end{pmatrix}\bar{\psi} + \begin{pmatrix} 0 \\ \left(-\sqrt[4]{N}\bar{e}\iota_p + \varepsilon_{aq}\right) \\ \varepsilon_{bq} \end{pmatrix}\right. \\
&\quad \left. - \begin{pmatrix} A_f(\theta(e))_1'A_q \\ \left(\frac{1}{h_N(\theta_{0,N})}G_f(\theta(e))'A_q\right) \\ 0 \end{pmatrix}\bar{\psi} - \begin{pmatrix} 0 \\ \left(V_{aq,abd}B(N)\right) \\ V_{bq,abd}B(N) \end{pmatrix}\right.
\end{aligned}
$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$\times \left(B(N)'V_{abd}B(N)\right)^{-1}\left(e^2\sigma^2\begin{pmatrix}\iota_p\\0\end{pmatrix}+B(N)'\begin{pmatrix}\varepsilon_a\\\varepsilon_b\\\varepsilon_d\end{pmatrix}\right)\Bigg\}+o_p(1),$$

$$=\begin{pmatrix}0\\-\begin{pmatrix}\iota_p\\0\end{pmatrix}\bar{e}\end{pmatrix}+\frac{1}{\sqrt[4]{N}}\begin{pmatrix}0\\\nu\end{pmatrix}+o_p(N^{-1/4})$$

$$=-\begin{pmatrix}0\\\begin{pmatrix}\iota_p\\0\end{pmatrix}\end{pmatrix}\bar{e}+o_p(1),$$

where we have rescaled, since all the higher-order terms have dropped out, and which shows that the additional components in Theorem 1* compared to Theorem 1 do not affect the limit behavior of $\hat{D}_N(\theta(e))$ up to order $N^{-1/4}$. The specification of ν is:

$$\nu=-\left(\begin{pmatrix}V_{aq,abd}B(N)\\V_{bq,abd}B(N)\end{pmatrix}\right)\left(B(N)'V_{abd}B(N)\right)^{-1}\begin{pmatrix}\iota_p\\0\end{pmatrix}e^2\sigma^2$$
$$\times\left[\begin{pmatrix}\varepsilon_{aq}\\\varepsilon_{bq}\end{pmatrix}-\begin{pmatrix}V_{aq,abd}B(N)\\V_{bq,abd}B(N)\end{pmatrix}\left(B(N)'V_{abd}B(N)\right)^{-1}B(N)'\begin{pmatrix}\varepsilon_a\\\varepsilon_b\\\varepsilon_d\end{pmatrix}\right],$$

which is independent of the limit behavior of $\sqrt{N}g_{f,T}(\theta(e))$.

We obtain the limit behavior of $\sqrt{N}\hat{D}_N(\theta(e))'\hat{V}_{ff}(\theta(e))^{-1}D_N(\theta(e))$ from:

$$\sqrt{N}\hat{D}_N(\theta(e))'\hat{V}_{ff}(\theta(e))^{-1}\hat{D}_N(\theta(e))$$
$$=\left[\sqrt[4]{N}(h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp)'\hat{D}_N(\theta(e))\right]'$$
$$\times((h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp))^{-1}$$
$$\times\left[\sqrt[4]{N}(h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp)'\hat{D}_N(\theta(e))\right]$$
$$=\left[\begin{pmatrix}\iota_p\\0\end{pmatrix}\bar{e}+\frac{1}{\sqrt[4]{N}}\nu\right]'\left(B(N)'V_{abd}B(N)\right)^{-1}\left[\begin{pmatrix}\iota_p\\0\end{pmatrix}\bar{e}+\frac{1}{\sqrt[4]{N}}\nu\right]+o_p(1)$$
$$=\bar{e}^2\begin{pmatrix}\iota_p\\0\end{pmatrix}'\left(B(N)'V_{abd}B(N)\right)^{-1}\begin{pmatrix}\iota_p\\0\end{pmatrix}+o_p(1)$$

and

$$N^{\frac{3}{4}}\hat{D}_N(\theta(e))'\hat{V}_{ff}(\theta(e))^{-1}f_N(\theta(e))$$
$$=\left[\sqrt[4]{N}(h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp)'\hat{D}_N(\theta(e))\right]'$$
$$\times((h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp)'\hat{V}_{ff}(\theta(e))(h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp))^{-1}$$
$$\times\sqrt{N}\left[(h_N(\theta_{0,N})A_f(\theta(e))_1\vdots A_f(\theta(e))_\perp)'f_N(\theta(e))\right]$$
$$=\left[\left(B(N)'V_{abd}B(N)\right)^{-\frac{1}{2}\prime}\left[\begin{pmatrix}\iota_p\\0\end{pmatrix}\bar{e}+\frac{1}{\sqrt[4]{N}}\nu\right]\right]'$$

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

$$\times \left(B(N)'V_{abd}B(N)\right)^{-\frac{1}{2}} \left(e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) + o_p(1)$$

$$= \bar{e} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} \left(e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) + o_p(1).$$

Upon combining the behavior of the above two components, we obtain the large sample behavior of the KLM statistic to test $H_p : \theta(e) = 1 + \frac{e}{\sqrt[4]{N}}$ under Assumptions 1 and 2(a), which can for samples of (large) size N be specified as:

$$\begin{aligned}
&\text{KLM}(\theta(e)) \\
&= N f_N(\theta(e))' \hat{V}_{ff}(\theta(e))^{-1} \hat{D}_N(\theta(e)) \left[\hat{D}_N(\theta(e))' \hat{V}_{ff}(\theta(e))^{-1} \hat{D}_N(\theta(e)) \right]^{-1} \\
&\quad \times \hat{D}_N(\theta(e))' \hat{V}_{ff}(\theta(e))^{-1} f_N(\theta(e)) \\
&= \left[N^{\frac{3}{4}} \hat{D}_N(\theta(e))' \hat{V}_{ff}(\theta(e))^{-1} f_N(\theta(e)) \right]' \left[\sqrt{N} \hat{D}_N(\theta(e))' \hat{V}_{ff}(\theta(e))^{-1} \hat{D}_N(\theta(e)) \right]^{-1} \\
&\left[N^{\frac{3}{4}} \hat{D}_N(\theta(e))' \hat{V}_{ff}(\theta(e))^{-1} f_N(\theta(e)) \right] \\
&= \left(e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right)' \left(B(N)'V_{abd}B(N)\right)^{-1} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \bar{e} \\
&\quad \times \left[\bar{e}^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \right]^{-1} \\
&\quad \times \bar{e} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} \left(e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \\
&= [\kappa + \eta]'[\kappa + \eta] + o_p(1) \\
&\sim \chi^2(\delta(N), 1) + o_p(1),
\end{aligned}$$

where $\bar{e}$ cancels out, since it is a scalar, $\kappa = \left(\begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \right)^{\frac{1}{2}} e^2\sigma^2$, $\eta = \left(\begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \right)^{-\frac{1}{2}} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \sim N(0,1)$,

and

$$\delta(N) = (e\sigma)^4 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \left(B(N)'V_{abd}B(N)\right)^{-1} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}$$

on the right-hand side of the above specification depends on N, which is important to obtain an accurate approximation because of the quartic root convergence rates.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

REFERENCES

Ahn, S. C. & P. Schmidt (1995) Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5–27.

Ahn, S.C. & G.M. Thomas (2006) Likehood Based Inference for Dynamic Panel Data Models. Unpublished manuscript.

Alvarez, J. & M. Arellano (2004) Robust Likelihood Estimation of Dynamic Panel Data Models. CEMFI Working Paper No. 0421.

Anderson, T. W. & C. Hsiao (1981) Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598–606.

Anderson, T. W. & C. Hsiao (1982) Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18, 47–82.

Anderson, T. W. & H. Rubin (1949) Estimation of the parameters of a single equation in a complete set of stochastic equations. *The Annals of Mathematical Statistics* 21, 570–582.

Andrews, D. W. K., M. J. Moreira & J. H. Stock (2006) Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.

Andrews, I. (2016) Conditional linear combination tests for weakly identified models. *Econometrica* 84, 2155–2182.

Andrews, I. & A. Mikusheva (2016) Conditional inference with a functional nuisance parameter. *Econometrica* 84, 1571–1612.

Arellano, M. & S. Bond (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277–297.

Arellano, M. & O. Bover (1995) Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29–51.

Blundell, R. & S. Bond (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.

Bond, S., C. Nauges, & F. Windmeijer (2005) Unit Roots: Identification and Testing in Micro Panels. CEMMAP Working paper CWP07/05, Centre for microdata methods and practice, University College, London.

Bond, S. & F. Windmeijer (2005) Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews* 24, 1–37.

Bun, M. J. G. & F. Windmeijer (2010) The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13, 95–126.

Davidson, R. & J.G. MacKinnon (2002) Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* 66, 1–26.

Dhaene, G. & K. Jochmans (2016) Likelihood inference in an autoregression with fixed effects. *Econometric Theory* 31, 1178–1215.

Dovonon, P. & A. R. Hall (2018) The asymptotic properties of GMM and indirect inference under second-order identification. *Journal of Econometrics* 205, 76–111.

Dovonon, P., A. R. Hall & F. Kleibergen (2020) Inference in second-order identified models. *Journal of Econometrics* 218, 346–372.

Dovonon, P. & E. Renault (2013) Testing for common conditionally heteroskedastic factors. *Econometrica* 81, 2561–2586.

Dufour, J. M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1388.

Dufour, J. M. & M. Taamouti (2005) Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* 73, 1351–1365.

Engle, R. F. & C. W. J. Granger (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.

Guggenberger, P., F. Kleibergen & S. Mavroeidis (2019) A more powerful Anderson–Rubin test in linear instrumental variables regression. *Quantitative Economics* 10, 487–526.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press

Guggenberger, P., F. Kleibergen, S. Mavroeidis & L. Chen (2012) On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression. *Econometrica* 80, 2649–2666.

Hahn, J., J. Hausman & G. Kuersteiner (2007) Long difference instrumental variable estimation for dynamic panel models with fixed effects. *Journal of Econometrics* 140, 574–617.

Han, C. & P. C. B. Phillips (2010) GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26, 119–151.

Hansen, L. P. (1982) Large sample properties of generalized method moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L. P., J. Heaton & A. Yaron (1996) Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14, 262–280.

Hsiao, C., M. H. Pesaran & A. K. Tahmiscioglu (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107–150.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Kleibergen, F. (2005) Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103–1124.

Kleibergen, F. (2021) Efficient size correct subset inference in homoskedastic linear instrumental variables regression. *Journal of Econometrics* 221, 78–96.

Kleibergen, F., L. Kong, & Z. Zhan (2020) Identification Robust Testing of Risk Premia in Finite Samples. Journal of Financial Econometrics, Forthcoming.

Kleibergen, F. & Z. Zhan (2020) Robust inference for consumption-based asset pricing. *Journal of Finance* 75, 507–550.

Kruiniger, H. (2002) On the Estimation of Panel Regression Models with Fixed Effects. Manuscript, Queen Mary University.

Kruiniger, H. (2009) GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory* 25, 1348–1391.

Kruiniger, H. (2013) Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. *Journal of econometrics* 173, 175–188.

Madsen, E. (2003) GMM Estimators and Unit Root Tests in the AR(1) Panel Data Model. Centre for Applied Micro Econometrics Working paper 2003-11, University of Copenhagen.

Moreira, M. J. (2003) A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.

Newey, W. K. & F. Windmeijer (2009) Generalized method of moments with many weak moment conditions. *Econometrica* 77, 687–719.

Nickell, S. J. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1426.

Phillips, P. C. B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Phillips, P. C. B. (2018) Dynamic panel Anderson–Hsiao estimation with roots near unity. *Econometric Theory* 34, 253–276.

Staiger, D. & J. H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J. H. & J. H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.

https://doi.org/10.1017/S026646662100027X Published online by Cambridge University Press